PATENT

VANM243.1APC1

OIPE JC10 FEB 1 1 2005 PATENT & TRADEMARK OFFICE

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant	:	Gabant *et al.*
Appl. No.	:	10/031,021
Filed	:	March 19, 2002
For	:	NON-HUMAN GENETICALLY MODIFIED MAMMAL LACKING THE ALPHA-FETOPROTEIN
Examiner	:	Nguyen, D. T.
Group Art Unit	:	1632

DECLARATION UNDER 37 C.F.R. §1.132

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

1. This Declaration is being submitted to demonstrate that the knowledge of how to create a modification in a mouse AFP gene using ES cell technology was readily available to a person skilled in the art at the time the US Provisional Application No.: 60/143,269 to which this Application claims priority was filed.

2. I am an inventor on the above-identified patent application and am familiar with the specification and prosecution history.

3. I have extensive experience in the field of the claimed invention as indicated in the attached Curriculum Vitae provided herewith as Exhibit A.

4. In some embodiments of the presently claimed mice, at least one allele of the endogenous genetic sequence of the wild-type alpha-fetoprotein (AFP) is disrupted to prevent the mice from making sufficient active AFP to confer fertility. When such a genetically modified mouse has just one allele of the AFP gene modified, it can be mated with another such mouse to obtain

female mice which lack an active AFP gene these animals were shown to be infertile. We conducted experiments where most of the sequences of exon 1, exon 2 and exon 3 of AFP were deleted. The mating of mice heterozygous for such mutations resulted in offspring homozygous for that deletion which did not express any functional AFP (functional knock-out). The female littermates of the offspring were found to be sterile. Using the same strategy, a skilled artisan could delete all the exons of AFP resulting in complete loss of expression (complete genetic knock-out).

Furthermore, other strategies for modifying the AFP gene to generate female mice which do not produce sufficient active AFP protein to confer fertility existed by July 12, 1999, the date to which the present application claims priority. For example, mutations resulting in a frame-shift would cause loss of expression of functional AFP. Such mutations can be of any length, including point-mutations, partial deletions or complete deletions. One of skill in the art would readily be able to design such mutations as of July 12, 1999, since the sequences of the mouse AFP gene and protein were known as of July 1999 see Gorin and Tilghman 1980 PNAS, 77: 1351-1355; W. Law and Dugaiczyk 1981 Nature, 291: 201; Eiferman et al., 1981 Nature 294: 713-718.

Exhibit 1, Askew et al. 1993 *Molecular and Cell Biology*, 13:4115-4121, describes how to introduce point mutations into a protein in mouse embryonic stem cells using "tag-and-exchange" strategy. The authors successfully altered two amino acids in α2 isoform Na,K-ATPase to make the enzyme resistant to inhibition by cardiac glycosides. The same strategy would work if the goal was to introduce a frame-shift mutation in the AFP gene, resulting in loss of expression of a functional protein.

Exhibit 2, Rubinstein et al. 1993 *Nucleic Acid Research*, 21:2613-2617, describes introduction of a point mutation into the mouse genome by homologous recombination in ES cells using a replacement type vector with a selectable marker. The authors successfully introduced a single nucleotide insertion into the mouse proopiomelanocortin gene (POMC) to transcriptionally silence the gene.

Appl. No. : 10/031,021
Filed : March 19. 2002

Exhibit 3, Valancius et al. 1991 *Molecular and Cell Biology*, **11**:1402-1408, describes an "in-out" targeting procedure to introduce an insertion into a mouse gene of a mouse ES cell line. The authors successfully corrected the mutant hypoxanthine phosphoribosyltransferase gene (*hprt*) by introducing a 4-bp insert, causing expression of functional *HPRT*$^{+}$.

Exhibits 4-8 (Hasty et al. 1991 *Molecular and Cell Biology*, **11**:5586-5591; Hasty et al. 1991 *Molecular and Cell Biology*, **11**:1509-4517; Bradley et al. 1998 *Int. J. Dev. Biol.* 42:943-950; WO 95/23862 and WO 99/09150) further show that by July 12, 1999 the technology of disrupting genes in ES cells thereby creating functional knock-outs was well-developed, if tedious and labor-intensive.

Therefore, I assert that as of July 12, 1999, using the disclosure of the present application, one skilled in the art could readily make a genetically modified mouse comprising a mutation, a partial deletion or a total deletion in the endogenous genetic sequence of alpha-fetoprotein which prevents the expression of sufficient active alpha-fetoprotein to confer fertility.

7. I declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful, false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or patent issuing therefrom.

Dated: 4 of February 05

By: Philippe Gabant, PhD

O:\DOCS\MXG\MXG-6226.DOC
102804

Exhibit A

GABANT Philippe, PhD

Functions

Head of the Department de Génie Génétique DGG of BioVallée

CEO/co-founder of Delphi Genetics SA

Nationality: Belgian

Date and place of birth: 7 January 1965, Brussels, Belgium

Address: rue Van Bemmel 9, 1210 Bruxelles, Belgium

Phone: (office) +32 (0)71378525
e-mail: philippe.gabant@biovallee.be
pgabant@delphigenetics.com

Qualifications

BioVallée

Since September 2001: Head of the DGG « Département de Génie Génétique ». The DGG is a 15 persons research team (4 post docs, 1 MD and a technical staff of 10 persons). The DGG develop innovative tools in DNA engineering, Genome engineering, genotyping and phenotyping.

Université Libre de Bruxelles

September 1997-September 2001: Laboratoire de Biologie du Développement on the characterisation of knock-out mice. Development of a "mouse targeting facility" at the Institut de Biologie et Médecine Moléculaire (IBMM-ULB). Development of new cloning tools for the construction of mouse model.

Edinburgh University

September 1996 - September 1997: Centre for Genome Research (CGR) of the University of Edinburgh, Laboratory of Professor Lesley Forrester. The main topic of this research was the determination of the alpha-fetoprotein (AFP) physiological function by a gene targeting approach. I also studied a gene trapped integration expressed in the foetal liver.

HEC Saint-Louis

Master in management (1995): "License en management de l'innovation technologique"

Université Libre de Bruxelles

December 1993 - September 1996: Laboratoire de Biologie du Développement of Professors Josiane and Claude Szpirer in the Département de Biologie Moléculaire of the Université Libre de Bruxelles.

PhD thesis: October 1988 - November 1993: Laboratoire de Génétique of Pr. René Thomas under the direction of Pr. Martine Thilly Couturier.Title of the thesis: "Analyse des mécanismes de maintenance des plasmides des groupes d'incompatibilités IncHI et IncA/C". (Analysis of the maintenance mechanism of plasmids belonging to the IncHI and IncA/C groups) (la plus grande distinction).

Bachelor of Zoological Sciences: September 1988 (grande distinction)

Languages

Mother langage: French
Foreign langages: English, Dutch (school knowledge)

Scientific collaborations

Prof Kenn Gerdes, SDU, Odense, Denmark: analysis of poison-antidote function and applications
Prof Michel Milinkovith, IBMM,-ULB, Belgium : analysis of poison-antidote function and applications
Prof Ramon Diaz: CSIC, Madrid, Spain: analysis of poison-antidote function and applications

Prof. Bruno André: IBMM-ULB, Belgium, Laboratoire de Physiologie Cellulaire: Targeting of the mouse *RhcG* gene.

Prof. Oberdan Leo: IBMM-ULB, Belgium, Laboratoire de physiologie animale: Targeting of the mouse *pbef* gene.

Prof. Jacques Mahillon: UCL, Belgium, Laboratoire de Microbiologie: Analysis of bacterial poison/antidote systems.

Prof. Henry Alexandre: Université Mons-Hainaut, UMH-Faculté de Médecine, Belgium. The topic of our collaboration is the phenotypic characterisation of *Afp* knock-out mice.

Dr. Johan Smitz: AZ-VUB, Belgium, Radioimmunology and Reproductive Biology, Belgium. The topic of our collaboration is the phenotypic characterisation of *Afp* knock-out mice.

Dr. Jennifer Nichols: Centre for Genome Research, University of Edinburgh, UK. The topic of our collaboration is the phenotypic characterisation of *Afp* knock-out mice.

Dr. Catherine Llanes: Laboratoire de Bactériologie, Université de Franche-Comté, France. The topic of the collaboration was the isolation and the molecular analysis of a replicon specific of the IncA/C plasmids.

Prof. Diane Taylor: I spent one month (December 1992) in the Laboratory of Microbiology, University of Edmonton, Canada. The research topic of the collaboration was the maintenance of the IncHI1 plasmids.

International advanced scientific courses

- University of Amsterdam, Netherlands, course entitled : « Strategic protein purification » from the 19 to 23 of June 1995.
- Participation on the advanced FEBS course entitled: "Tools for visualizing molecular interactions controlling gene expression" at the Uppsala University, Sweden, 8-17 August 1994.

International advanced management courses

- « Licensing and negociation Seminar » Brussels, Belgium, 22 and 23 June 2000.
- « The business platform : a model for detecting critical sucess for high-tech start-ups » Paris, France, 26 January 2001.

Meetings

More than 25 Poster presentation and oral communications

Publications

Gabant Philippe, Newnham Peter, Taylor Diane and Couturier Martine. Isolation and location on the R27 map of two replicons and an incompatibility determinant specific for the IncHI1 plasmids. *Journal of Bacteriology* December 1993, Vol. 175, N°.23, p7697-7701.

Gabant Philippe, Ouazzani Chahdi Abdeljawad and Couturier Martine. Nucleotide sequence and replication characteristics of RepHI1B a replicon specific of the IncHI1 plasmids. *Plasmid* 1994, Vol. 31, p111-120.

Llanes Catherine, Gabant Philippe, Couturier Martine and Yvon Michel-Briand. Cloning and characterization of the IncA/C plasmid RA1 replicon. *Journal of Bacteriology* June 1994, Vol. 176 N° 11, p3403-3407.

Bernard Philippe, Gabant Philippe, El Bahasi Moustapha and Couturier Martine. Use of the ccdB killer gene for construction of positive selection vectors.*Gene* 1994, Vol 148, p71-74.

Llanes Catherine, Gabant Philippe, Couturier Martine, Laurence Bayer and Patrick Plesiat. Molecular analysis of the broad-host-range RepA/C replicon. *Plasmid* 1996, Vol. 36, p26-35.

Marie-France Henriette, Gabant Philippe, Pierre-Luc Drèze, Claude Szpirer and Josiane Szpirer. Negative regulation of the alpha-foetoprotein gene in fibroblast: identification and characterization of *cis* and *trans* elements. *Folia Biologica* 1997 Vol 42, p319-327.

Ouazzani Chahdi Abdeljawad, Gabant Philippe and Couturier Martine. Molecular analysis of RepHI1B: A replicon specific of the IncHI1 plasmids. *Molecular General Genetics* 1997 Vol 255, p477-486.

Gabant Philippe, Drèze Pierre-Luc, Van Reeth Thierry, Szpirer Josiane and Szpirer Claude. Use of bifunctional *lacZα-ccdB* genes for selective cloning of PCR products. BioTechniques 1997 Vol 23, p938-941.

Gabant Philippe, Cédric Szpirer, Couturier Martine and Faelen Michel. New positive-selection vectors adapted to the genetic analysis of various Gram-negative bacteria and their plasmids. *Gene* 1998 Vol 207, p87-92

Van Reeth Thierry, Drèze Pierre-Luc, Szpirer Josiane, Szpirer Claude and <u>Gabant Philippe.</u> Positive selection vectors to generate fused genes for the expression of His tagged proteins. *BioTechniques* 1998 Vol. 25, p898-904.

Szpirer Claude, Tissir Fadel, Riviere Michèle, Van Vooren Pascale, Kela Johanna, Lallemand Françoise, <u>Gabant Philippe</u>, Hoebee B. , Klinga-Levan, Levan Goran, Szpirer Josiane. Rat Chromosome 2: assignment of the genes encoding cyclin B1, interleukin 6 signal transducer, and proprotein convertase 1 to the Mcs1-containing region and identification of new microsatellite markers. *Mammalian Genome* 1999 Vol. 10, p30-34.

<u>P. Gabant</u>, T. Van Reeth, P-L Drèze, M. Faelen, C. Szpirer, J. Szpirer. A new positive selection system based on the *parD* (*kis/kid*) system of the R1 plasmid, *Biotechniques* 2000 Vol. 28, p784-788.

Scohy S, <u>Gabant P</u>, Szpirer C, Szpirer J. Identification of an enhancer and an alternative promoter in the first intron of the alpha-fetoprotein gene. *Nucleic Acids Res* 2000 Vol 28, p3743-3751.

S. Scohy, <u>P. Gabant</u>, T. Van Reeth, V. Hertveldt, P-L Drèze, P. Van Vooren, M. Rivière, J. Szpirer and C. Szpirer. Identification of KLF13 and KLF14 (SP6), novel members of the SP/XKLF transcription factor family *Genomics* 2000 Vol 70, p93-101.

<u>P. Gabant</u>, T. Van Reeth, P-L Drèze, M. Faelen, C. Szpirer, J. Szpirer. Use of "poison-antidote" systems for the selection of recombinants. *Plasmid* 2001

Thierry Van Reeth, Pierre-Luc. Drèze, Josiane Szpirer, Claude Szpirer and <u>Philippe Gabant.</u> Positive selection vectors to generate fused genes for the expression of His tagged proteins. In "Cloning vectors for protein expression" Eds. 2001

Laurence Van Melderen, Cédric Szpirer, <u>Philippe Gabant</u>. Les systèmes poison-antidotes de leurs compréhension à leurs applications en biotechnologie. *BioFutur* 2001 n°211 p 46-p49.

<u>Philippe Gabant</u>, Cédric Szpirer, Laurence Van Melderen. Plasmid poison/antidote systems: their functions and technological applications. In "Advance plasmid research" Eds. 2001

<u>Philippe Gabant</u>, Lesley Forrester, Jennifer. Nichols, Thierry Van Reeth, Christelle De Mees, Alistair Watt, Johan Smitz, Henri Alexandre, Claude Szpirer and Josiane Szpirer Alpha-fetoprotein the major fetal serum protein is not required for development but is essential for female fertility. *PNAS 2002* Vol. 99, p 12865-12870

Exhibit 1

MOLECULAR AND CELLULAR BIOLOGY, July 1993, p. 4115–4124 Vol. 13, No. 7
0270-7306/93/074115-10$02.00/0

Site-Directed Point Mutations in Embryonic Stem Cells: a Gene-Targeting Tag-and-Exchange Strategy

G. ROGER ASKEW, THOMAS DOETSCHMAN, AND JERRY B. LINGREL*

Department of Molecular Genetics, Biochemistry, and Microbiology, University of Cincinnati College of Medicine, 231 Bethesda Avenue, Cincinnati, Ohio 45267-0524

Received 25 January 1993/Returned for modification 9 March 1993/Accepted 1 April 1993

Sequential gene targeting was used to introduce point mutations into one α2 isoform Na,K-ATPase homolog in mouse embryonic stem (ES) cells. In the first round of targeted replacement, the gene was tagged with selectable markers by insertion of a Neo^r/HSV-tk gene cassette, and this event was selected for by gain of neomycin (G418) resistance. In the second targeted replacement event, the tagged genomic sequence was exchanged with a vector consisting of homologous genomic sequences carrying five site-directed nucleotide substitutions. Embryonic stem cell clones modified by exchange with the mutation vector were selected for loss of the HSV-tk gene by resistance to ganciclovir. Candidate clones were further screened and identified by polymerase chain reaction and Southern blot analysis. By this strategy, the endogenous α2 isoform Na,K-ATPase gene was altered to encode two other amino acids so that the enzyme is resistant to inhibition by cardiac glycosides while maintaining its transmembrane ion-pumping function. Since the initial tagging event and the subsequent mutation-exchange event are independent of one another, a tagged cell line can be used to generate a variety of mutant lines by exchange with various mutation vectors at the tagged locus. This method should be useful for testing specific mutations introduced into the genomes of tissue culture cells and animals and for developing animal models encompassing the mutational variability of known genetic disorders.

The concerted application of targeted gene modification and production of animals derived from embryonic stem (ES) cells has established a powerful method for studying gene function in the context of the whole animal (for reviews, see references 6 and 13). In this scheme, modification of a particular locus is performed first by targeted homologous recombination (21) in cultured ES cells. In turn, genetically modified ES cells are used to generate mice by injection into blastocysts and allowing the chimeric blastocyst to mature to term following transfer into a pseudopregnant foster mother (7). This combination of techniques has been used to develop many lines of mice carrying null alleles at specific loci as a means of testing the general physiological roles of particular gene products. Such experiments provide an alternative method of studying planned mutations in known genes, rather than relying on fortuitous identification of random mutations. This approach has already led to new insights on gene function. In many cases the phenotypes of animals harboring gene-targeted null mutations have confirmed predictions which were based on knowledge of the gene product (14, 23, 28). However, in several instances, surprising phenotypic results have implied new functions and expanded the biological understanding of previously well-characterized proteins (20, 22).

This relatively new tool of mammalian genetics presents the potential for testing the physiological significance of subtle functional mutations and for developing animal models of genetic disorders. To date, however, animals derived from gene-targeted ES cells have been designed exclusively to study the effect of null or "knockout" mutations. In order to examine the effect of more modest mutations which modify rather than eliminate the gene product, reliable methods for introducing such mutations into the genome, with fidelity, must be devised. In order to avoid artifacts due to extraneous genetic elements introduced during gene targeting, such as selectable markers, introduction of test mutations must be subtle in terms of perturbing the context of the gene. Since introduction of subtle mutations at most loci is not directly selectable, indirect selection for such recombination events is required.

Several strategies for introduction of minor mutations into nonselectable genes in ES cells have been explored. Microinjection of a mutation-targeting vector into ES cells resulted in targeted mutation of the *Hox* 1.1 gene at a frequency high enough to preclude the use of selectable markers (29). While the rate of targeted recombination was high, no reports of subsequent repetition of these experiments or of animals derived from the mutant ES cells have been published. The use of coelectroporation of a mutation vector and a selectable marker vector resulted in identification of targeted mutations of the hypoxanthine phosphoribosyltransferase (HPRT) gene (3). This method allows targeted mutation of loci which may be unable to express selectable marker genes. However, the rate of simultaneous homologous recombination and cotransfection with a selectable marker has been shown to be low (3, 17), and selection of cotransfected cells often results in multimeric insertions at the site of homologous recombination (3). A strategy for subtle mutation based on gene-targeted insertion (O-type recombination) and subsequent excision by intrachromosomal recombination, referred to as hit-and-run or in-out, has recently been reported (9, 25). Here we report the novel introduction of subtle point mutations into the α2 Na,K-ATPase gene of mouse ES cells by a sequential gene-targeted replacement (omega-type recombination) strategy which we call tag and exchange.

The target gene encodes one of the three isoforms of the catalytic (alpha) subunit of the ATP-dependent sodium-potassium pump. The Na,K-ATPase maintains the opposed sodium and potassium gradients across the plasma membranes of higher eukaryotes and also serves as the pharma-

* Corresponding author.

cological receptor for the cardiac glycoside (digitalis) class of drugs which act through inhibition of this enzyme. Point mutations introduced into the α2 Na,K-ATPase gene encode two amino acid changes which reduce the affinity of the enzyme for these drug ligands by 10^2- to 10^3-fold without interrupting the ion-pumping function of this integral membrane protein.

MATERIALS AND METHODS

Reagents. Enzymes used for cloning and analysis of DNA were purchased from Bethesda Research Laboratories, New England Biolabs, Pharmacia, Promega, Stratagene, and United States Biochemical. G418 (Geneticin) was purchased from Bethesda Research Laboratories, mitomycin was purchased from Boehringer Mannheim, and ganciclovir (GANC) sodium (Cytovene, manufactured for Syntex) was purchased locally. Oligonucleotides were synthesized on an Applied Biosystems 394 DNA synthesizer.

Cell culture, transfection, and selection of drug-resistant clones. D3 mouse embryonic stem cells were maintained by coculture on a feeder layer of mouse embryonic fibroblasts (which were arrested from proliferation by mitomycin treatment) in Dulbecco's modified Eagle's medium, 15% fetal bovine serum (Cellect-Gold; Flow Laboratories), and 10^{-4} M β-mercaptoethanol (Sigma). Mouse embryonic fibroblasts were prepared from 16-day postcoital embryos, which were transgenic for the neomycin resistance gene, according to the 4°C trypsinization method (2). For transfection, ES cells were dispersed by trypsin digestion for 2 to 3 min at 37°C, washed, resuspended in ES cell culture medium at 2×10^7 cells per ml, and electroporated with 3.5 nM vector DNA at 200 μF and 400 V/cm. All targeting vectors were excised as inserts from their parent plasmid and gel purified prior to transfection. Electroporated ES cells were seeded onto neomycin-resistant embryonic fibroblasts at 5×10^5 cells per 100-mm dish. These electroporation conditions resulted in 10 to 25% ES cell viability.

Neomycin (G418)-resistant ES cell clones were selected by exposure to 400 μg of G418 per ml starting 24 h after electroporation. The concentration of G418 was 400 μg/ml for the first 24 h of selection and was reduced to 300 μg/ml for the remainder of the selection period. During selection, cells were fed every other day until resistant colonies were picked on days 13 and 14 following electroporation. G418^r homologous recombinants were maintained under G418 selection (300 μg/ml) prior to transfection with the mutation vector.

Ganciclovir-resistant ES cell clones were selected by exposure to 2.5 μM Cytovene for 4 days, at which time colonies were picked and transferred with a disposable-tip pipetter to individual wells of 24-well tissue culture dishes preseeded with embryonic fibroblasts.

In order to minimize the time in culture for isolated ES cell clones, each line was expanded by a single passage and an aliquot was frozen for storage. Two days after transfer to individual wells, ES cells were dispersed with trypsin and allowed to resettle in the same well. Two to three days later, each well was roughly confluent and the cells were trypsinized and resuspended in 250 μl of freezing medium (culture medium plus 10% dimethyl sulfoxide). One drop (about 10 μl) of this cell suspension was used to seed each of two sets of 24-well tissue culture dishes: one set for G418 selection and one set for maintaining the lines and for preparation of genomic DNA. The remaining (230 μl) of cell suspension was frozen for storage in 500-μl microcentrifuge tubes (Marsh Biomedical). Using this scheme, we froze all selected GANC^r cell lines within 14 days of electroporation and maintained the lines in culture for screening and identification of homologous recombinants.

Neo^r/HSV-tk expression cassettes. All constructs were prepared in the pBluescript II SK (+) cloning vector (Stratagene). For both cassettes the Neo^r gene was first subcloned directly from pMC1Neo (Stratagene) into pBluescript II SK (+) as an *Xho*I-*Sal*I fragment, generating pBSSK-Neo. The thymidine kinase gene of herpes simplex virus (HSV-tk) driven by the pMC1Neo (*Xho*I-*Eco*RI) promoter (a gift from Jay Degen) was then subcloned into the *Cla*I site of pBSSK-Neo by using *Cla*I linkers. The head-to-tail and tail-to-tail constructs were isolated from the same ligation reactions.

Preparation and analysis of genomic DNA. Small-scale preparations of genomic DNA were prepared from ES cell clones grown to confluence on 24-well plates by the method of Laird et al. (12), with the exception that cell lysates were transferred to microcentrifuge tubes for isopropanol precipitation. Large-scale genomic DNA isolation from confluent ES cell cultures on P-100 plates was done by standard proteinase K digestion and phenol-chloroform extraction followed by RNase treatment. ES cell clones from the first round of gene targeting were screened for homologous recombination by polymerase chain reaction (PCR) with a primer specific to the 3′ end of the HSV-tk gene (of the sequence 5′-CTGAGCAGACAGACCCATGC-3′) and a primer specific to intron 10 of the α2 isoform Na,K-ATPase gene (of the sequence 5′-CACGATTGCAGAGACCAGCG-3′).

GANC^r G418^s ES cell clones were screened for the presence of site-directed mutations in genomic DNA by PCR amplification and diagnostic restriction digestion of a 270-bp sequence encompassing the nucleotide substitutions of the mutation vector. This 270-bp fragment was amplified with a forward primer of the sequence 5′-AGCTCAGGACATTC TGG-3′ based on the exon 4 sequence and a reverse primer of the sequence 5′-TGAGCTCCTGACTGGGTGTG-3′ based on the intron 4 sequence. The genomic PCR products were extracted once with chloroform; one half of the product was then digested for 2 h with *Bst*BI restriction endonuclease, and the other half was used as a nondigested control. Control and *Bst*BI-digested PCR products were separated on 3% agarose gels and visualized by ethidium bromide staining. In this assay, the 270-bp PCR product of the site-directed mutants is uniquely cleaved by *Bst*BI, generating diagnostic 176- and 94-bp digestion fragments, while the PCR product of the wild-type allele remains at 270 bp.

To check for delivery of each nucleotide substitution in targeted mutations of homologous recombinants, the 270-bp PCR product (described above) was cloned and sequenced. This 270-bp fragment was cloned by blunt-end ligation into the *Eco*RV site of pBluescript II SK (Stratagene) after blunting of the PCR product with T4 DNA polymerase. Sequence of the cloned PCR product was determined by double-strand sequencing with Sequenase (United States Biochemicals).

Genomic Southern blot analysis was performed according to the standard protocol. Briefly, 10 μg of genomic DNA was digested overnight in a 300-μl reaction volume, ethanol precipitated, and electrophoreticaly separated on a 0.8% agarose gel. DNA was transferred to Magna N/T membrane (MSI) by capillary action with 10× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate). After transfer, DNA was fixed to the membrane by baking and prehybridized in standard buffer. Southern blots were hybridized with either

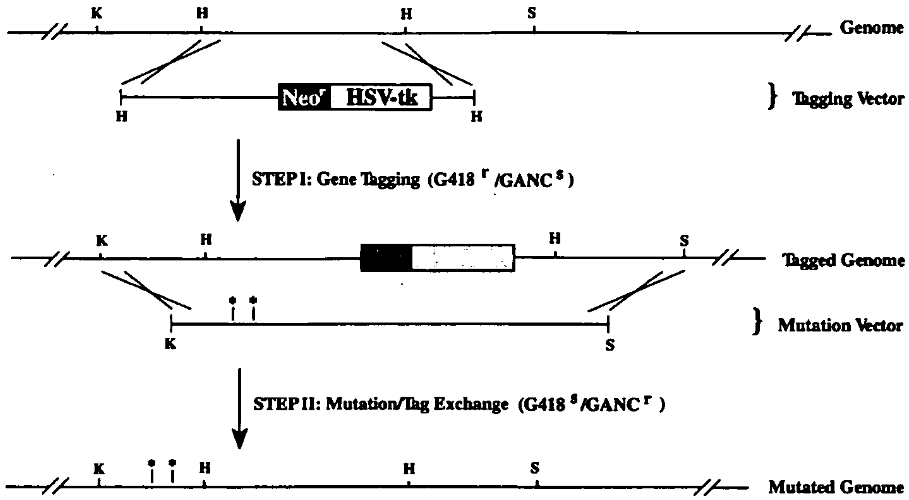


FIG. 1. Schematic diagram illustrating the gene tag-and-exchange strategy. In this illustration, the target site is the region between the *Hin*dIII (H) sites. The tagging vector is composed of a Neo^r/HSV-tk selectable marker cassette flanked by roughly 3 and 1 kb (5' and 3', respectively) of genomic DNA homologous to the target sequence. Step I represents the homologous recombination and subsequent tagging of the target gene with the tagging vector. The mutation vector is composed of a larger fragment of genomic DNA homologous to the target sequence and contains sequence mutations (**). Step II represents homologous recombination between the tagged locus and the mutation vector resulting in a mutated genomic sequence in a marker-free context. The *Kpn*I (K) and *Sma*I (S) site are included to provide landmarks.

random hexamer-labeled probe in standard hybridization buffer (50% formamide, 6× SSC, 5× Denhardt's solution, 10% dextran sulfate, 100 μg of denatured salmon sperm DNA per ml, 0.1% sodium dodecyl sulfate [SDS]) or with an end-labeled oligonucleotide probe in oligonucleotide hybridization buffer (5× SSC, 1% SDS, 5× Denhardt's solution, 100 μg of denatured salmon sperm DNA per ml).

Preparation and analysis of DNA clones. A mouse strain 129 sv genomic DNA library of *Sau*3AI partial-digest fragments (18 to 21 kb) was prepared in the Lambda Dash II vector (Stratagene). Screening this library with a radiolabeled rat cDNA probe specific for the α2 isoform Na,K-ATPase gene resulted in isolation of 14 clones containing fragments of the α2 isoform Na,K-ATPase gene. Two of these clones, λ41-2 and λ78-1, which overlap, represent the entire α2 isoform gene and were characterized by mapping the recognition sites for several restriction enzymes by indirect end labeling of partial digests. The general exon organization of the gene was then mapped by Southern blot analysis with serial fragments of the sheep α1 Na,K-ATPase cDNA. Genomic inserts were positively identified as α2 isoform Na,K-ATPase gene fragments by partial sequencing of exons 4 and 8 to 10.

Targeting and control vectors, described in Results, were prepared with subcloned restriction fragments from the genomic insert of clone λ78-1. The mutation vector was prepared with an 8.6-kb *Kpn*I-*Sma*I genomic fragment of the λ78-1 clone. Site-directed mutations of coding sequence within exon 4 were introduced by oligonucleotide-directed mutagenesis by the method of Kunkel (11). Briefly, a 1.1-kb *Bam*HI fragment of the 8.6-kb *Kpn*I-*Sma*I subclone was transferred to M13mp18 to generate M13mp18-α21.1 and a uracil-rich template of M13mp18-α21.1 was prepared from BW313 cells. A mutagenesis oligonucleotide with the sequence 5'-TACGGTATCCGGGCCGCCATGGAGGACGA GCCTTCGAATGATGATGTGAGCC-3' was phosphorylated and used to prime DNA synthesis from the U-rich template. Inserts of M13mp18-α21.1 clones containing the mutated sequence were sequenced in their entirety, and the mutant 1.1-kb *Bam*HI fragment was transferred back into the pmα2-8.6 kb K-S subclone, generating the mutation construct pmα2-8.6RD.

RESULTS

Tag-and-exchange strategy and testing the positive and negative selection cassette. The strategy for introduction of point mutations into the α2 isoform Na,K-ATPase gene in a marker-free context was to sequentially replace the gene sequences by homologous recombination. The general scheme used, shown in Fig. 1, was to first tag the gene of interest with selectable markers and then replace the tagged sequence with a sequence carrying a cluster of site-directed point mutations. This method requires two selectable markers, a positive marker for identification of recombinants in the first targeting step and a negative marker for selection of recombinants from the second targeting step in which the tagged sequence is lost by exchange with the mutation vector. In the experiments described here, we used the neomycin resistance gene (Neo^r) and the HSV-tk gene, which confers sensitivity to ganciclovir, as positive and negative selection markers, respectively. These marker genes were fused to prepare a positive and negative selection cassette in which the genes are driven by identical promoters.

Two positive and negative selection constructs were prepared by fusing the Neo^r and HSV-tk genes and inserting the pair into a polylinker, providing a cassette which could be easily subcloned. Each gene is driven independently by the HSV-tk promoter fused to polyomavirus enhancer repeats, which together were subcloned as a 177-bp *Xho*I-*Eco*RI fragment from pMC1Neo (Stratagene). In planning the construction of these selection cassettes, all four possible orientations of the two genes were considered. Two orientations were chosen for construction and empirically tested: one with the genes head to tail with the Neo^r gene 5' of the

HSV-tk gene, and the other with the genes oriented tail to tail.

The selection properties of the head-to-tail and tail-to-tail selection cassettes were examined in order to determine which provided G418 resistance accompanied by the lowest reversion to GANC^r. To measure reversion from a GANC^s to a GANC^r phenotype, ES cells were first electroporated independently with either the head-to-tail or tail-to-tail marker cassette, and G418^r recombinants were selected in 300 μg of G418 per ml. For each construct, cells from a plate of 300 to 400 G418^r colonies were examined for their plating efficiency (reversion to GANC^r) in 2.5 μM GANC and 300 μg of G418 per ml. The results of these empirical measurements demonstrated that the head-to-tail configuration provides a lower rate of reversion than the tail-to-tail, 0.03 versus 0.4% plating efficiency, respectively, in ganciclovir. This translates to an estimated background of 300 colonies for a typical gene-targeting electroporation experiment in which 10^7 cells are electroporated under conditions resulting in 90% cell death. The head-to-tail cassette was therefore used as the selectable tag for construction of the gene-tagging vector. While these measurements were useful for choosing which positive and negative marker cassette to use in vector construction, it should be emphasized that the results indicate reversion at random sites of integration and may be subject to locus-specific rates of HSV-tk inactivation.

Targeting constructs. Both targeting constructs were prepared with cloned genomic sequences from the same mouse strain (129) as that used in preparation of the D3 line of blastocyst-derived ES cells (5). This precaution was taken to maximize the degree of sequence homology between the targeting vector and the target gene, thereby avoiding potential reduction of targeting efficiency due to genomic sequence polymorphisms between vector and target DNA. This concern is supported by recent comparative studies of targeting frequency between homologous and heterologous vector and target combinations in ES cells (4, 18, 26). The gene-tagging vector (mα2-4Ntk) illustrated in Fig. 2 was constructed by inserting the 3-kb Neo^r/HSV-tk cassette at the *Bal*I site of exon 8 in a 4-kb genomic *Hin*dIII fragment of the α2 isoform Na,K-ATPase gene. Genomic sequences flanking the selection cassette are 2.9 kb 5′ and 1.1 kb 3′ of the positive and negative selection cassette, and the marker genes are in the same transcriptional orientation as the α2 isoform Na,K-ATPase gene.

A positive control construct (mα2-1.8Ntk) illustrated in Fig. 2 was also prepared for optimizing PCR conditions used to screen for homologous recombination in the first round of gene targeting. The PCR assay for homologous recombination between the tagging vector and the target α2 isoform Na,K-ATPase locus requires a set of primers which only participate in amplification of a DNA template unique to the tagged α2 isoform Na,K-ATPase gene. This primer set includes a forward primer specific for the HSV-tk gene of the tagging vector and a reverse primer specific for the genomic target sequence immediately flanking the 3′ end of the region of homology between vector and target. The positive control construct (Fig. 2) was prepared by inserting the selection cassette into the *Bal*I site of exon 8 in a 1.8-kb *Eco*RI fragment of α2 isoform Na,K-ATPase gene. This construct mimics the product of homologous recombination between the tagging vector and the chromosomal target in that it contains a template for PCR which is identical to the correctly tagged α2 locus. It was used to optimize and check for successful PCR.

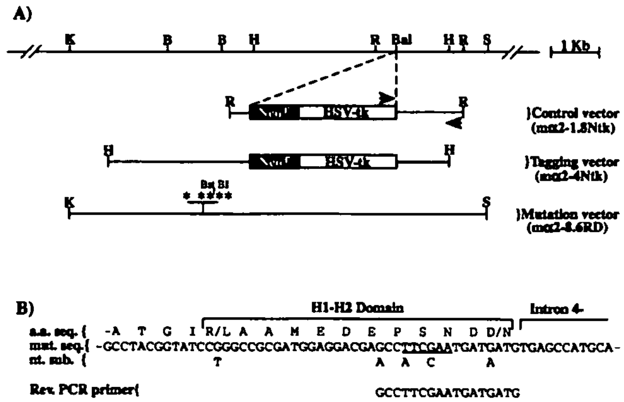


FIG. 2. Gene-targeting vectors. (A) Schematic diagram of the genomic restriction map of the α2 isoform Na,K-ATPase gene. The restriction sites represented include *Bal*I (Bal), *Bam*HI (B), *Eco*RI (R), *Kpn*I (K), and *Sma*I (S). The control vector (mα2-1.8Ntk) was constructed by insertion of the head-to-tail Neo^r/HSV-tk marker cassette into the *Bal*I site of the 1.8-kb *Eco*RI fragment of the α2 isoform Na,K-ATPase gene. The horizontal arrows represent the forward (leftmost) and reverse (rightmost) primers used to screen for homologous recombinants by PCR. Note that the reverse primer is unique to the genome and to the control construct. The tagging vector (mα2-4Ntk) was constructed by insertion of the selection cassette into the *Bal*I site of a 4-kb *Hin*dIII genomic fragment of the gene. The mutation vector (mα2-8.6RD) is composed of an 8.6-kb *Kpn*I-*Sma*I genomic fragment of the α2 isoform Na,K-ATPase gene and carries five point mutations (* ****) and a unique *Bst*BI recognition site in exon 4. (B) Details of the nucleotide substitutions in exon 4 of the mutation vector. The amino acid sequence (a.a. seq.) encoded by the mutation shows the substitutions (L111R and N122D) at the border residues of the extracellular H1-H2 domain. The mutation sequence (mut. seq.) shows the *Bst*BI recognition sequence (underlined) resulting from A→T and C→G substitutions. The wild-type nucleotides which were substituted (nt. sub.) are in the line below the mutation sequence. A reverse primer (Rev. PCR primer) designed for mutation-specific PCR amplification is also shown.

The mutation vector (mα2-8.6RD) illustrated in Fig. 2 consists of an 8.6-kb genomic *Kpn*I-*Sma*I fragment of the α2 gene carrying five nucleotide substitutions in exon 4. The objectives of these planned point mutations were twofold. First and foremost are the codon changes resulting in two amino acid substitutions which modify the cardiac glycoside receptor function of the mouse α2 isoform Na,K-ATPase. Second, this combination of mutations provides several means of screening and identifying cell clones which have been mutated by exchange at the tagged locus. The point mutations were first introduced by M13-mediated site-directed mutagenesis (11) into a 1.1-kb *Bam*HI fragment, which was then switched with its wild-type counterpart to complete the mutation vector. The nucleotide and amino acid substitutions encoded by the mutation vector are shown in Fig. 2B. All five nucleotide substitutions are within the coding region for the first extracellular domain of the enzyme. The amino acid substitutions of the border residues, L111R and N122D, of this extracellular domain were previously demonstrated to confer a 10^3- to 10^4-fold resistance to enzyme inhibition by cardiac glycosides by decreasing the affinity of the enzyme for this steroid ligand (15). To provide a means of screening and identifying correctly mutated ES cell clones, three additional nucleotide substitutions were introduced at wobble positions. These mutations introduce a

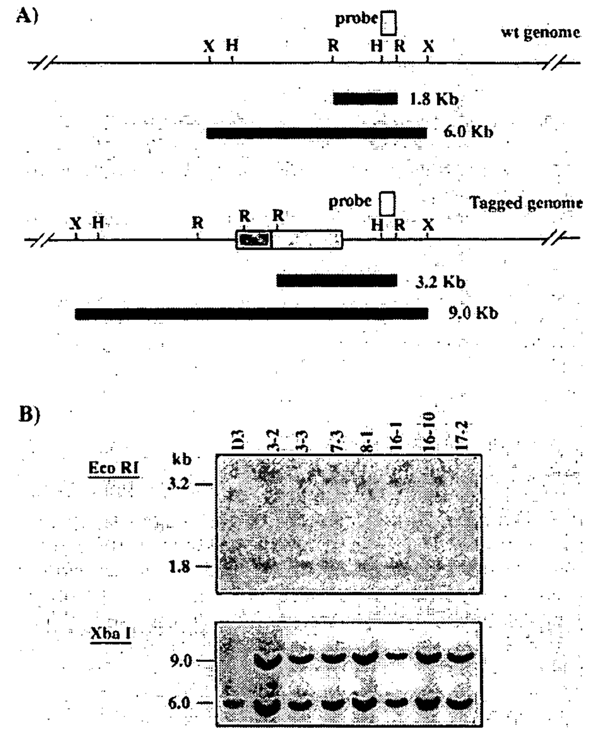


FIG. 3. Analysis of the tagging event. (A) Restriction maps of the wild-type (wt) and tagged alleles used to predict the diagnostic restriction fragment lengths of *Eco*RI (R)- and *Xba*I (X)-digested genomic DNA. The probe (empty block) is a 308-bp *Hin*dIII-*Eco*RI fragment which is external to the genomic sequence of the tagging vector. The predicted *Eco*RI and *Xba*I fragments for the wild-type and tagged α2 isoform Na,K-ATPase alleles which hybridize to the 308-bp probe are drawn as solid bars below the restriction maps. (B) Southern blot analysis of *Xba*I- and *Eco*RI-digested genomic DNA of D3 ES cells and seven G418^r cell lines tagged by homologous recombination with the tagging vector.

unique restriction endonuclease recognition site and provide a mutation-specific primer sequence for PCR without altering the amino acid coding sequence. A *Bst*BI recognition site, unique to the mutated sequence, was introduced by the substitutions A→T in the proline 118 codon and C→G in the serine 119 codon. The wobble mutation A→G in the glutamate 117 codon adds a fourth nucleotide difference to a 14-nucleotide stretch in the mutated sequence of exon 4. This additional substitution was included to increase specificity for mutation-specific PCR with a primer which overlaps the divergent 14-nucleotide stretch.

Tagging the α2 isoform Na,K-ATPase gene. In the first gene-targeting event, an α2 isoform Na,K-ATPase homolog was tagged by replacement with the tagging vector (mα2-4Ntk) in which exon 8 had been disrupted by insertion of the Neo^r/HSV-tk selection cassette. The tagging vector was prepared as a gel-purified *Hin*dIII fragment and transfected into ES cells by electroporation. Transfectants were subject to G418 selection, and 600 G418^r clones were screened for homologous recombinants by PCR as described in Materials and Methods. PCR-positive clones were isolated, expanded, and analyzed by genomic Southern blotting for correct gene-targeted recombination according to diagnostic restriction endonuclease patterns. Identification of ES cell clones correctly targeted with the gene-tagging construct is illustrated in Fig. 3. Southern blots of genomic DNA isolated from PCR-positive clones, digested with either *Eco*RI or *Xba*I, were probed with a 308-bp genomic *Hin*dIII-*Eco*RI fragment which lies 3′ and adjacent to the flanking genomic sequences of the targeting construct. For *Eco*RI digests, this probe hybridizes to a 1.8-kb fragment of the wild-type α2 isoform Na,K-ATPase allele and a 3.2-kb fragment of a correctly targeted allele. For *Xba*I digests, the wild-type allele yields a 6-kb hybrid and the targeted allele yields a 9-kb hybrid with this probe. By these criteria, we isolated 15 ES cell lines which were tagged at the α2 isoform Na,K-ATPase locus with an overall targeting efficiency of 1:40 homologous recombinants to G418^r colonies (Table 1).

Exchange with the mutation vector. An ES cell line (16-10) with a tagged α2 isoform Na,K-ATPase gene, as described above, was expanded and maintained under G418 selection. This cell line was then transfected by electroporation with the mutation vector (mα2-8.6RD) which was prepared as a gel-purified *Kpn*I-*Sma*I fragment. Transfected cells were selected for resistance to 2.5 μM ganciclovir beginning at day 0, 1, 2, 3, 4, or 5 following electroporation. This variable lag period prior to selection was designed to test for the optimum recovery period required for decay and dilution of HSV-tk activity, since recombinants which have lost the HSV-tk gene will remain sensitive to ganciclovir as long as active enzyme exists in the cell. Selection was complete after 3 days, and resistant colonies were isolated and transferred with a disposable-tip pipetter to 24-well plates for expansion. The electroporation of 10^7 cells yielded an estimated total of 1,000 colonies which survived the GANC selection for loss of HSV-tk activity, and 500 of these clones were isolated for further analysis.

Upon expansion of the isolated GANC^r clones, a portion of the cells was frozen for storage and the remaining cells were passaged in duplicate to new 24-well culture plates. One set of plates was used for preparation of genomic DNA, and the other set was used to analyze the clones for the loss of the Neo^r gene by G418 selection. Clones carrying the exchange-mediated mutation should have lost the entire Neo^r/HSV-tk cassette and therefore would be G418^s

TABLE 1. Comparative targeting frequency data for tag and exchange at the α2 Na,K-ATPase locus

Event	No. of cells: Electroporated (10^7)	No. of cells: Surviving electroporation (10^6)	Total no. of colonies	No. of clones picked and screened	No. of targeted cell lines	Frequency: Per cell	Frequency: Per resistant colony
Tagging	1	1	600 (G418^r)	600	15	1.5×10^{-5}	1:40
Exchange	1	2.5	1,000 (GANC^r)	216 (days 0–3), 288 (days 4–5)	0 (days 0–3), 4 (days 4–5)	5.5×10^{-6} (days 4–5)	1:54 (days 4–5)

TABLE 2. Selection[a] data for tag and exchange at the α2 Na,K-ATPase locus

Day of GANC selection (postelectroporation)	No. of GANC[r] colonies		No. of GANC[r] G418[s] colonies	Targeted exchange
	Per plate	Picked and screened		
0	17	17	2	0
1	32	26	6	0
2	29	29	1	0
3	34	144	9	0
4	31	144	17	3
5	32	144	13	1
Subtotal (days 4 and 5)	31	288	30	4
Total		504	48	4

[a] Selection for exchange by homologous recombination.

GANC[r]. To test for the loss of the Neo[r] gene, one set of the isolated clones was selected in 400 μg of G418 per ml for 3 weeks. Of the 504 GANC[r] clones, only 48 were also G418[s], indicating that at least 90% of the GANC[r] clones had not lost the selection cassette and therefore could not be complete products of homologous recombination with the mutation vector. The GANC and G418 selection data are provided in Table 2. The 48 Neo[r−]/HSV-tk[−] clones were subsequently examined for the presence of the mutation vector by PCR, and then homologous recombinants were verified by genomic Southern blot analysis.

Testing for the presence of the mutation vector in GANC[r] G418[s] clones provided an independent screen for clonal products of targeted exchange. Because the probability of random recombination with the vector and concomitant loss of the selectable markers in the same cell is exceedingly low, clones which have gained the vector and lost the markers are most likely homologous recombinants (see Discussion). The GANC[r] G418[s] clones were examined for the presence of the mutation vector by PCR amplification of a 270-bp fragment encompassing the point mutations and testing for the mutation-specific *Bst*BI cleavage of this 270-bp PCR product (data not shown). This assay revealed the presence of the mutated sequence in four clones (6-20, 19-15, 20-3, and 20-23) which were then analyzed for homologous recombination as described below.

The genomic location of the mutant sequence in these four candidate clones was established by Southern blot analysis. To distinguish between random integration and homologous recombination, genomic Southern blots of the candidate clones were screened with mutant and wild-type oligonucleotide probes, as illustrated in Fig. 4. Wild-type and mutant versions of the 52-base oligonucleotide used to introduce the point mutations proved to be specific probes for this analysis. The conditions for specificity were determined with cloned sequences and then applied to genomic Southern blot analysis. Southern blots of *Eco*RV-, *Kpn*I-, and *Sma*I-digested genomic DNA were hybridized with the probe which can distinguish between the wild-type and mutated α2 sequences. These restriction digests were chosen because they represent a combination of infrequent recognition sites, providing a diagnostic set of predicted restriction fragment patterns. The predicted *Eco*RV fragment length is a result of one internal and one external recognition site. Since the mutation vector was transfected as a gel-purified *Kpn*I-*Sma*I fragment, the *Kpn*I and *Sma*I sites are disrupted and therefore are not transmitted with the vector. If the mutation vector has been correctly targeted to the α2 isoform Na,K-ATPase gene, then the mutant probe and the wild-type probe should hybridize to identical patterns of restriction fragments and fragments unique to the tagged allele should be absent. A nontargeted integration of the mutation vector would likely result in an alternate pattern of restriction fragments, generating a hybridization pattern different from that of the wild type.

The mutation oligonucleotide probe generated the predicted (wild-type) hybridization pattern in the absence of additional bands one would expect of nontargeted integration (Fig. 4B). This probe hybridizes to a 6.6-kb *Eco*RV fragment, a 5.0-kb *Kpn*I-*Eco*RV fragment, and a 16-kb *Kpn*I-*Sma*I fragment of genomic DNA from clones 6-20, 19-15, 20-3, and 20-23, indicating that the mutation construct has been correctly targeted to one α2 isoform Na,K-ATPase gene in these clones. In addition, the *Kpn*I-*Sma*I-digested DNA displays a band which is unique to the tagged cell line and distinguishes the tagged from the wild-type α2 Na,K-ATPase allele. In the bottom panel of Fig. 4B, conversion of the 9.3-kb *Kpn*I-*Sma*I fragment of the tagged allele to the 16-kb *Kpn*I-*Sma*I fragment of the nontagged allele further demonstrates exchange of the tagged sequence with the mutated sequence. These results demonstrate homologous recombination and effectively rule out the possibility of random integration of the vector in the four candidate clones.

To complete the identification of targeted clones 6-15, 19-20, 20-3, and 20-23, the region encompassing the five nucleotide substitutions was cloned from each cell line and sequenced. A 270-bp genomic fragment spanning the mutated region was amplified by PCR and cloned into a plasmid vector. PCR clones of the mutant allele were identified by *Bst*BI sensitivity and sequenced. These sequences (Fig. 4C) demonstrate delivery of all five point mutations to each of the modified cell lines.

All four correctly modified ES cell lines were from a population of 288 colonies which were allowed to grow for 4 or 5 days posttransfection before being selected in GANC. Of the 144 GANC[r] colonies selected beginning on day 3 posttransfection, no correctly modified cells were identified. Therefore, 4 days is sufficient, if not the minimal required time, for dilution and decay of residual HSV-tk activity to allow for selection of gene-targeted exchange of the selection cassette. The selection results for this experiment are provided in Table 2.

Analysis of the GANC[r] background population from the exchange step. The background population of GANC[r] ES cell colonies selected from the targeted exchange step is derived from cells which have lost HSV-tk activity without acquiring the desired mutation by homologous recombination between the tagged locus and the mutation vector. Two subpopulations make up this background population: one (GANC[r] G418[r]) which has lost activity of HSV-tk only, and the other (GANC[r] G418[s]) which has lost activity of both HSV-tk and Neo[r]. There are two basic explanations for this loss of marker activity; one is the physical loss of the marker gene(s), and the other is inactivation of marker gene expression. As a general assessment of the origin of the GANC[r] background, the loss versus inactivation of the selectable markers was determined by assaying for the presence or absence of the Neo[r]/HSV-tk selection cassette in GANC[r] clones. In these genomic Southern blot assays of *Xba*I-digested DNA, a probe for the α2 Na,K-ATPase gene reveals a 9-kb fragment if the marker genes are present and a 6-kb fragment of the wild-type allele if the marker genes are

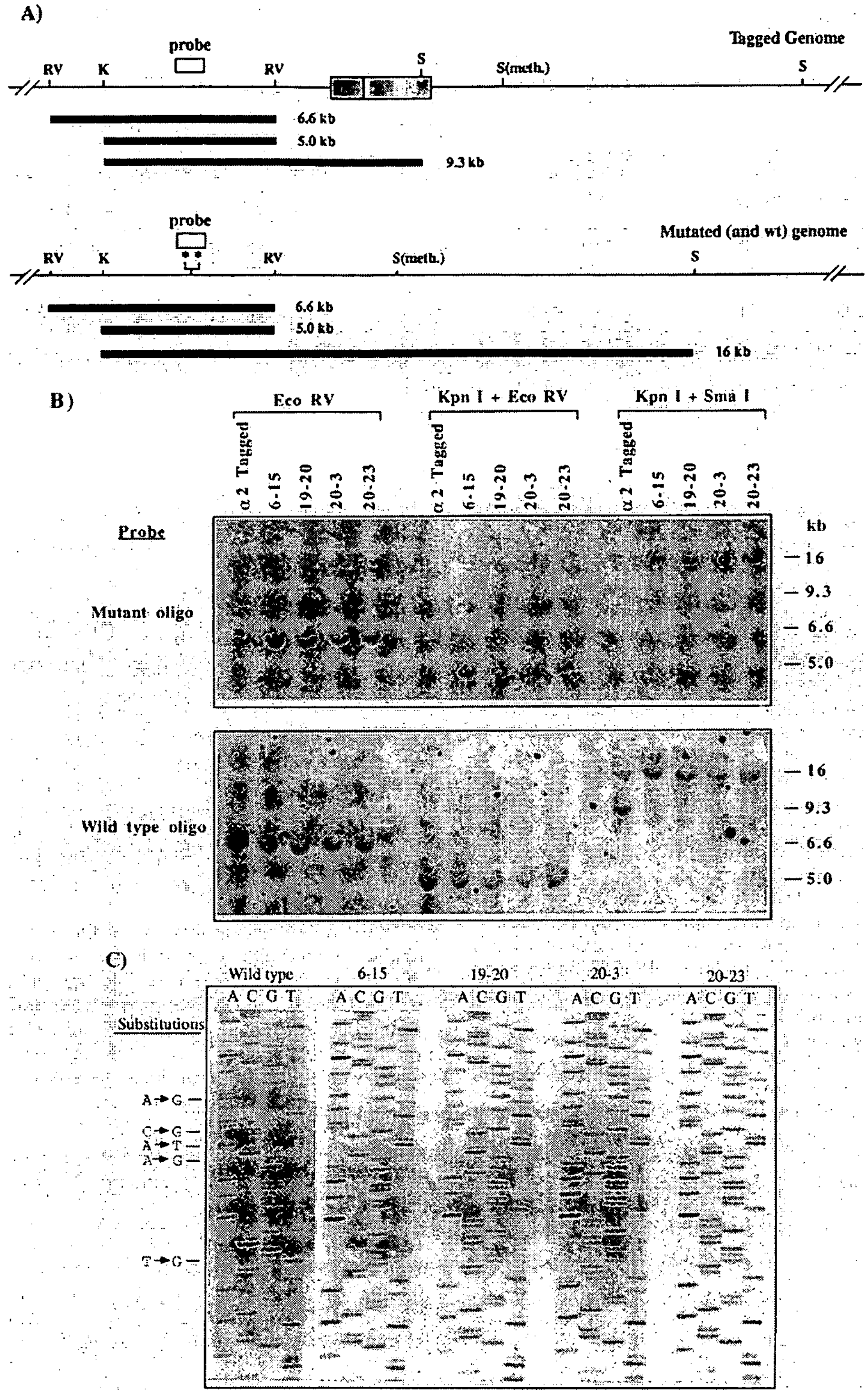


FIG. 4. Analysis of the exchange event. (A) Restriction maps of the tagged allele (top) and the mutated (exchanged) allele (bottom). The sites for cleavage by *Eco*RV (RV), *Kpn*I (K), and *Sma*I (S) were used to predict diagnostic restriction fragment lengths. A *Sma*I site [labeled S(meth.)], identified in genomic clones, is insensitive to *Sma*I digestion in any of the genomic analyses. The predicted fragment lengths for the tagged and mutated alleles are illustrated by solid bars below the maps. The wild-type (wt) and mutant probes (empty block) were 52-bp oligonucleotides encompassing the region of nucleotide substitutions and were used to distinguish between the wild-type and mutant sequences. (B) Southern blot analysis showing the tagged parent cell line (α2 tagged) and the four cell lines mutated by homologous recombination between the mutation vector and the tagged allele. The lower panel probed with the wild-type oligonucleotide shows the hybridization pattern of the wild-type and tagged alleles. The upper panel represents the same blot probed with the mutant oligonucleotide, demonstrating the hybridization pattern of the modified allele in each of the four modified cell lines. Note that since the mutant oligonucleotide (top panel) is not 100% free of cross-hybridization, the faint bands of cross-hybridization in the wild-type (α2 Tagged) lanes provide size markers with which to compare the hybridization patterns of the mutant sequence. (C) Sequence analysis showing all five nucleotide substitutions (indicated at the left) delivered to the modified α2 Na,K-ATPase gene.

TABLE 3. Frequencies for presence and absence of the Neo^r/HSV-tk sequences in GANC^r background clones as scored by Southern blot analysis

Phenotype	No. of clones		Neo^r/HSV-tk cassette[a] (no. present/no. absent)	Selectable activity		Disposition (%) of selectable marker genes		
	Selected	Assayed		HSV-tk	Neo^r	Single inactivation (HSV-tk only)	Double inactivation (HSV-tk and Neo^r)	Absent
GANC^r G418^r	456	50	50/0	–	+	100 (50/50[b])	0 (0/50[b])	0 (0/50[b])
GANC^r G418^s	44	44	4/40	–	–	0 (0/44[b])	9 (4/44[b])	91 (40/44[b])
Total	500	94	460 (456[c] + 4)/40			91 (456[c]/500[d])	0.8 (4/500[d])	8 (40/500[d])

[a] Presence or absence of the Neo^r/HSV-tk cassette was determined by Southern blot analysis of *Xho*I-digested genomic DNA hybridized to an α2 Na,K-ATPase probe as for Fig. 4. In these assays the presence of the Neo^r/HSV-tk cassette is indicated by the 9-kb *Xho*I fragment of the tagged allele. Clones which have lost the HSV-tk cassette reveal only the 6-kb *Xho*I fragment of the wild-type allele.
[b] Number with marker gene/number of clones assayed.
[c] Extrapolated as 100% of the GANC^r G418^r clones.
[d] Number with marker gene/number of clones selected.

absent from the genome, as demonstrated in Fig. 3. In analyzing these data, the presence of the marker cassette coupled with loss of selectable activity was interpreted as marker inactivation.

Fifty of the GANC^r G418^r clones and all 44 GANC^r G418^s background clones were analyzed. No gross rearrangements were detected on the basis of fragment sizes in this analysis. The data in Table 3 show that 100% (50 of 50) of the GANC^r G418^r clones examined have maintained the marker cassette, indicating that this subpopulation of background clones is the product of inactivation of HSV-tk gene expression rather than physical loss of the gene. This inactivation must be relatively stable and may reflect, for example, methylation inactivation of transcription or mutation in the coding sequence of this marker gene. In contrast, 91% (40 of 44) of the GANC^r G418^s background clones have lost the marker cassette, while 9% (4 of 44) of these clones have maintained this cassette and are apparently the products of inactivation of both Neo^r and HSV-tk. While inactivation of Neo^r may be mechanistically similar to HSV-tk inactivation, loss of the marker cassette is probably the result of gene conversion, chromosome loss, or incomplete homologous recombination in which a crossover has occurred in the region between the marker cassette of the tagged allele and the point mutation of the mutations vector.

Assuming that the 50 GANC^r G418^r clones assayed accurately represent the entire GANC^r G418^r subpopulation, the origin of both GANC^r subpopulations can be expressed in relative terms (Table 3). Thus, inactivation of HSV-tk accounts for 92% (91% + 0.8%) of the GANC^r background, while double inactivation (both Neo^r and HSV-tk) accounts for only 0.8% of the total background. Physical loss of the marker gene cassette accounts for 8% (44 of 504) of the total background clones.

DISCUSSION

The experiments presented here demonstrate the novel use of sequential gene-targeted replacement for the subtle introduction of nucleotide substitutions in one homolog of the α2 isoform Na,K-ATPase gene of murine ES cells. These mutations include two codon changes which result in amino acid substitutions and three functionally silent point mutations designed for identification of correctly modified cells. Na,K-ATPase is the receptor for cardiac glycosides which inhibit enzymatic activity of this receptor and are used in the treatment of congestive heart failure and certain cardiomyopathies. The amino acid substitutions introduced reduce affinity for cardiac glycosides, thereby conferring resistance to enzymatic inhibition by these ligands (15). Endogenous cardiac glycosides have recently been reported as normal steroid components of plasma in several mammals including humans (8, 19). One of the long-term goals of these experiments is to reveal the physiological role of these putative endogenous cardiac glycosides by developing a strain of ES cell-derived mice expressing α2 isoform Na,K-ATPase with significantly reduced affinity for these endogenous ligands.

Theoretically, ES cells with a tagged allele could be used to generate animals carrying null mutations of the tagged gene, as long as the selectable cassette disrupted a coding region. However, Braun et al. (1) have found that expression of HSV-tk in postmeiotic germ cells disrupts maturation of sperm, resulting in male sterility. Thus, ES cells tagged with an HSV-tk gene are probably not useful for generating knockout mice.

Homologous recombination of the gene-tagging step and of the targeted exchange with the mutation vector occur at approximately the same rate. The frequency of tagging the α2 isoform Na,K-ATPase gene was 15:600 (tagged lines to Neo^r colonies or 1.5×10^{-5} per viable electroporated cell). In turn, the frequency of the exchange reaction at the tagged locus was 4:288 (mutant lines to GANC^r G418^s colonies [GANC added 4 days posttransfection] or 5.5×10^{-6} per viable electroporated cell). These values are well within the range of variability for targeting experiments and indicate that the targeting efficiencies for the tagging and exchange steps are roughly equivalent.

Several groups have tried the sequential replacement strategy of subtle gene modification. In one account, the first step was achieved by disruption of the cystic fibrosis transmembrane conductance regulator gene in ES cells with a Neo^r/HSV-tk cassette (16). Other, unpublished accounts of attempted tag-and-exchange indicate that in many cases reversion to homozygosity of the wild-type allele, to the exclusion of gene-targeted recombination, is the major (or only) genetic event producing cells selected for the loss of marker gene activity. This may be due to poor targeting frequency, high background in the exchange selection, and/or elimination of positive clones by premature application of selection.

Design of the tagging and mutation vectors should provide for the most efficient targeting of both vectors and most efficient selection of correctly modified clones. The targeting vectors should include a length of homologous sequence, 4 to 10 kb, suitable to provide high-frequency gene targeting

(4) with the selectable marker cassette located near the position of the planned mutation. The vector arms should also be at least 1 kb to maximize precise recombination (24, 27). There are two important considerations in positioning the mutations in the mutation vector: distance from the vector ends and relative proximity of the selection cassette. First, as observed in the careful study of Jiang et al. (10), it is important to place the desired mutation at least 400 bp, ideally 1 to 2 kb, from the ends of the targeting vector to optimize delivery of the mutant sequence to the genome. Second, the mutation(s) should be close to or overlap the selectable markers in the tagged allele. This precaution should optimize correct modification by minimizing incomplete homologous recombination in which a crossover occurring between the mutant sequence and the selectable markers results in exchange with the marker cassette or the mutation sequence but not both.

The tag-and-exchange strategy presented here relies on omega-type homologous recombination. Two groups have recently reported the introduction of a subtle mutations by using O-type homologous recombination followed by intrachromosomal recombination. Hasty et al. (9) used the Neo^r and HSV-tk selection scheme to mutate the HPRT and *Hox-2.6* genes, whereas Valancius and Smithies (25) used the HPRT gene marked with a 4-base insertion to modify the HPRT locus in ES cells. While Hasty et al. reported that the reversion rate to GANC^r was 3.8×10^{-3} per cell generation, their recovery rate of successfully mutated clones was not provided. The frequency of the "out reaction" in the Valancius and Smithies studies was 8×10^{-7} per cell plated. In the work presented here, the overall rate of reversion to GANC^r was $10^3/10^7$ cells electroporated. Considering the 25% postelectroporation survival rate, the reversion to GANC^r becomes 4×10^{-4} per viable cell. The efficiency of targeted exchange with the tagged sequence is 5.5×10^{-6} per viable cell.

Analysis of the GANC^r population of ES cell clones from the exchange step suggests a more efficient selection scheme for the exchange step. In our experiments, three kinds of genetic events which lead to generation of GANC^r cells in the exchange step occurred: (i) background loss of the Neo^r/HSV-tk cassette, (ii) inactivation of the HSV-tk gene, and (iii) replacement of the selection cassette with the mutation construct by homologous recombination. The background frequency of physical loss of the marker genes was 8%, and the frequency of HSV-tk inactivation was 92%, while only 0.8% of the clones were inactivated for both HSV-tk and Neo^r. Since inactivation of a single marker gene accounted for roughly 90% of the background, while loss or inactivation of both marker genes was roughly 10%, eliminating single inactivation as a source of background would theoretically provide a 10-fold enrichment for homologous recombinants. If two negative markers were used, providing double selection, for the exchange step, single inactivation would not contribute to the background. Therefore, a selection scheme including two independent negative selection genes, for example HPRT and HSV-tk, would improve the efficiency of selection for targeted exchange.

In addition, we would like to emphasize the screening for the presence of to the mutation vector, for instance, by PCR, provides a useful enrichment step towards identifying homologous recombinants of the exchange reaction. Since targeted recombination must result in gain of the mutation vector (vector-positive) accompanied by loss of the selectable markers (GANC^r and G418^s), screening for presence or absence of the vector simply allows elimination of vector-negative clones from the population of potential homologous recombinants. Also, if nontargeted integration and marker loss are independent events, then because of the low frequencies for reversion to GANC^r G418^s (roughly 10^{-4} per viable cell) and for gain of the transfected vector (generally, 10^{-3} per viable cell), the frequency of both events occurring independently in the same cell is very low (roughly 10^{-7} per viable cell). This translates to less than one cell per transfection experiment, and consequently most if not all GANC^r G418^s vector$^+$ clones from a single transfection would be targeted recombinants. In the studies presented here, 4 of the 48 GANC^r G418^s clones had recombined with the mutation vector, and all 4 were homologous recombinants.

There are two advantages of targeted tag and exchange over the hit-and-run or in-out strategy. One advantage is the ability to use PCR analysis for presence of the mutation vector as a screen for homologous recombinants at the exchange step. In the insertion and excision scheme, this assay is precluded by the presence of the mutated sequence in the product of the insertion step. A second advantage is the potential of using the initial gene-tagged cell line as a substrate for replacement by different mutation constructs. Since the selection process of the second round of targeted recombinants is straightforward, production of cell lines carrying different mutations at the same locus would be routine once the cell line tagged at the gene of interest was prepared. In contrast, hit and run or in out requires that each mutation is separately targeted to the gene of interest, and then the targeted cell line is expanded and selected for the excision event. Therefore, tag and exchange may be especially useful in situations where multiple mutants are desired at a particular locus in a marker-free and otherwise wild-type context.

ACKNOWLEDGMENTS

We thank Jay Degen for his gift of clone pBSSK-HSV-tk and Mary Tyler for her skilled assistance in the culturing of ES cells.

This work was supported by NIH grant HL 41496 and by AHA grant SW-92-01-I.

REFERENCES

1. **Braun, R. E., D. Lo, C. A. Pinkert, G. Widera, R. A. Flavell, R. D. Palmiter, and R. L. Brinster.** 1990. Infertility in male transgenic mice: disruption of sperm development by HSV-tk expression in postmeiotic germ cells. Biol. Reprod. 43:684–693.
2. **Cole, R. J., and J. Paul.** 1966. The effects of erythropoietin on haem synthesis in mouse yolk sac and cultured foetal liver cells. J. Embryol. Exp. Morphol. 15:245–260.
3. **Davis, A. C., M. Wims, and A. Bradley.** 1992. Investigation of coelectroporation as a method for introducing small mutations into embryonic stem cells. Mol. Cell. Biol. 12:2769–2776.
4. **Deng, C., and M. R. Capecchi.** 1992. Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. Mol. Cell. Biol. 12:3365–3371.
5. **Doetschman, T. C., H. Eistetter, M. Katz, W. Schmidt, and R. Kemler.** 1985. The *in vitro* development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. J. Embryol. Exp. Morphol. 87:27–45.
6. **Evans, M. J.** 1989. Potential for genetic manipulation of mammals. Mol. Biol. Med. 6:557–565.
7. **Gossler, A., T. Doetschman, R. Korn, E. Serfling, and R. Kemler.** 1986. Transgenesis by means of blastocyst-derived embryonic stem cell lines. Proc. Natl. Acad. Sci. USA 83:9065–9069.
8. **Hamlyn, J., M. Blaustein, S. Bova, D. DuCharme, D. Harris, F. Mandel, W. Mathews, and J. Ludens.** 1991. Identification and characterization of a ouabain-like compound from human

plasma. Proc. Natl. Acad. Sci. USA 88:6259–6263.

9. **Hasty, P., R. Ramirez-Solis, R. Krumlauf, and A. Bradley.** 1991. Introduction of a subtle mutation into the Hox-2.6 locus in embryonic stem cells. Nature (London) 350:243–246.

10. **Jiang, L., A. Connor, and M. J. Shulman.** 1992. Effects of mutation position on frequency of marker rescue by homologous recombination. Mol. Cell. Biol. 12:3609–3613.

11. **Kunkel, T. A.** 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc. Natl. Acad. Sci. USA 82:488–492.

12. **Laird, P. W., A. Zijderveld, K. Linders, A. M. Rudnicki, R. Jaenisch, and A. Berns.** 1991. Simplified mammalian DNA isolation procedure. Nucleic Acids Res. 19:4293–4294.

13. **Mansour, S. L.** 1990. Gene targeting in murine embryonic stem cells: introduction of specific alterations into the mammalian genome. Genet. Anal. Tech. Appl. 7:219–227.

14. **McMahon, A. P., and A. Bradley.** 1990. The Wnt-t (Int-1) proto-oncogene is required for development of a large region of the mouse brain. Cell 62:1073–1085.

15. **Price, E., and J. B. Lingrel.** 1988. Structure-function relationships in the Na,K-ATPase α subunit: site-directed mutagenesis of glutamine-111 to arginine and asparagine-122 to aspartic acid generates a ouabain-resistant enzyme. Biochemistry 27:8400–8408.

16. **Ratcliff, R., M. J. Evans, J. Doran, B. J. Wainwright, R. Williamson, and W. H. Colledge.** 1992. Disruption of the cystic fibrosis transmembrane conductance regulator gene in embryonic stem cells by gene targeting. Transgenic Res. 1:177–181.

17. **Reid, L. H., E. G. Shesely, H. Kim, and O. Smithies.** 1991. Cotransformation and gene targeting in mouse embryonic stem cells. Mol. Cell. Biol. 11:2769–2777.

18. **Riele, H. T., R. Maandag, and A. Berns.** 1992. Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs. Proc. Natl. Acad. Sci. USA 89:5128–5132.

19. **Shaikh, I., B. Lau, B. Siegfried, and R. Valdes.** 1991. Isolation of digoxin-like immunoreactive factors from mammalian adrenal cortex. J. Biol. Chem. 266:13672–13678.

20. **Shull, M. M., I. Ormsby, A. B. Kier, S. Pawlowski, R. J. Diebold, M. Yin, R. Allen, C. Sidman, G. Proetzel, D. Calvin, N. Annunziata, and T. Doetschman.** 1992. Targeted disruption of the murine growth factor-β1 gene results in multifocal inflammatory disease. Nature (London) 359:693–699.

21. **Smithies, O., R. Gregg, S. Boggs, M. Koralewski, and R. Kucherlapati.** 1985. Insertion of DNA sequences into the human chromosomal β-globin locus by homologous recombination. Nature (London) 317:230–234.

22. **Soriano, P., C. Montgomery, R. Geske, and A. Bradley.** 1991. Targeted disruption of the c-src proto-oncogene leads to osteopetrosis in mice. Cell 64:693–702.

23. **Thomas, K. R., and M. R. Capecchi.** 1990. Targeted disruption of the murine Int-1 proto-oncogene resulting in severe abnormalities in midbrain and cerebellar development. Nature (London) 346:847–850.

24. **Thomas, K. R., C. Deng, and M. R. Capecchi.** 1992. High-fidelity gene targeting in embryonic stem cells by using sequence replacement vectors. Mol. Cell. Biol. 12:2919–2923.

25. **Valancius, V., and O. Smithies.** 1991. Testing an "in-out" targeting procedure for making subtle genomic modifications in mouse embryonic stem cells. Mol. Cell. Biol. 11:1402–1408.

26. **van Deursen, J., and B. Wieringa.** 1992. Targeting of the creatine kinase M gene in embryonic stem cells using isogenic and nonisogenic vectors. Nucleic Acids Res. 20:3815–3820.

27. **Zheng, H., P. Hasty, M. A. Brenneman, M. Grompe, R. A. Gibbs, J. H. Wilson, and A. Bradley.** 1991. Fidelity of targeted recombination in human fibroblasts and murine embryonic stem cells. Proc. Natl. Acad. Sci. USA 88:8067–8071.

28. **Zijlstra, M., M. Bix, N. E. Simister, J. M. Loring, D. H. Raulet, and R. Jaenisch.** 1990. β2-Microglobulin deficient mice lack $CD4^{-}8^{+}$ cytolytic T cells. Nature (London) 344:742–746.

29. **Zimmer, A., and P. Gruss.** 1989. Production of chimaeric mice containing embryonic stem (ES) cells carrying a homoeobox *Hox* 1.1 allele mutated by homologous recombination. Nature (London) 338:150–153.

Exhibit 2

Nucleic Acids Research, 1993, Vol. 21, No. 11 ***2613–2617***

Introduction of a point mutation into the mouse genome by homologous recombination in embryonic stem cells using a replacement type vector with a selectable marker

Marcelo Rubinstein[1], Miguel A.Japón[1] and Malcolm J.Low[1,2,*]
[1]Vollum Institute for Advanced Biomedical Research and [2]Department of Biochemistry and Molecular Biology, Oregon Health Sciences University, Portland, OR 97201-3098, USA

Received March 5, 1993; Revised and Accepted April 29, 1993

ABSTRACT

The introduction of small mutations instead of null alleles into the mouse genome has broad applications to the study of protein structure-function relationships and the creation of animal models of human genetic diseases. To test a simple mutational strategy we designed a targeting vector for the mouse pro-opiomelanocortin (POMC) gene containing a single nucleotide insertion that converts the initial tyrosine codon of β-endorphin 1-31 to a premature translational termination codon and introduces a unique *HpaI* endonuclease restriction site. The targeting vector also contains a *neo* cassette immediately 3′ to the last POMC exon and a herpes simplex virus thymidine kinase cassette to allow positive and negative selection. Homologous recombination occurred at a frequency of 1/30 clones of electroporated embryonic stem cells selected in G418 and gancyclovir. 10/11 clones identified initially by a polymerase chain reaction (PCR) strategy had the predicted structure without evidence of concatemer formation by Southern blot analysis. We used a combination of *Hpa* I digestion of PCR amplified fragments and direct nucleotide sequencing to further confirm that the point mutation was retained in 9/10 clones. The POMC gene was transcriptionally silent in embryonic stem cells and the targeted allele was not activated by the downstream phosphoglycerate kinase-1 promoter that transcribed the *neo* gene. Under the electroporation conditions used, we have demonstrated that a point mutation can be introduced with high efficiency and precision into the POMC gene using a replacement type vector containing a retained selectable marker without affecting expression of the allele in the embryonic stem cells. A similar strategy may be useful for a wide range of genes.

INTRODUCTION

Gene targeting in murine embryonic stem (ES) cells by homologous recombination and the subsequent production of transgenic mice carrying specific gene mutations from the transfected totipotential ES cells has been a recently exploited technique to create mouse models of human genetic diseases (1–2) and to explore the role of specific proteins in mammalian development (3–6). Virtually all studies reported to date have used gene targeting vectors containing deletions of essential protein coding regions resulting in null alleles. Although useful for many purposes, null alleles may result in a lethal phenotype precluding a complete analysis of gene function. Furthermore, many recessive genetic diseases are caused by point mutations that result in hypofunctional or dysfunctional protein products rather than by complete loss of function mutations. Consequently, several strategies have been described to introduce small mutations into the genome. Two groups have described methods to introduce subtle mutations based on insertional type vectors containing selectable markers (7–8). The necessity of spontaneous intrachromosomal recombination to resolve the targeted gene locus with a relatively high rate of gene conversions to wild type decreases the final frequency of recombination and the resulting increased passage number of the selected cells may reduce the chances of later germ line transmission of the desired mutations. Coelectroporation of a targeting vector and an independent selectable marker involves a more complicated primary screening procedure than standard Southern hybridization analysis or polymerase chain reaction of the initially obtained clones (9). An alternative method involving direct screening without the use of a selectable marker has not been widely used and is technically demanding (10).

A class of proteins whose function has not been studied using targeted mutagenesis is the polypeptide hormone family. Neuropeptides are typically processed from larger prohormones by a series of specific postranslational endoproteolytic cleavage steps to yield the biologically active peptides. Pro-

* To whom correspondence should be addressed at: Vollum Institute for Advanced Biomedical Research, Oregon Health Sciences University L-474, 3181 SW Sam Jackson Park Road, Portland, OR 97201-3098, USA

opiomelanocortin (POMC) is a neuropeptide precursor that generates a large family of peptides including adrenocorticotropic hormone (ACTH), melanocyte stimulating hormones (MSH), and β-endorphin, each possesing distinct biological activities (11–12). An analysis of the phenotype resulting from a null allele for the POMC gene would be complicated by the diversity of peptides generated from the single precursor. POMC is therefore an excellent model for the introduction of subtle mutations whose aim is to selectively alter the peptide profile resulting from prohormone processing. In this report we describe the design of a novel replacement type targeting vector including selectable markers that introduces a point mutation into the translated portion of the mouse POMC gene that will result in the selective lack of β-endorphin production from a carboxyl-terminally truncated precursor. The targeting vector recombined as predicted and the point mutation was retained with high efficiency in ES cells. An identical strategy should be applicable to the mutation of neuropeptide precursors at dibasic processing sites precluding the release of specific peptides from otherwize normal prohormones.

MATERIALS AND METHODS

Construction of the targeting vector

The plasmid pHAL (gift of M. Uhler), carrying a 10.2 kb Eco RI fragment containing the entire POMC gene isolated from a Balb/c mouse genomic library (13), was double digested with *Asp* 718 and *Hind* III and a 2.9 kb fragment was subcloned in pSelect (Promega, WI). A single base insertion was introduced by site-directed mutagenesis converting the codon for the amino-terminal tyrosine of b-endorphin into a termination codon (TAC to TAA). The same mutation creates a *Hpa* I restriction endonuclease site (GTTAC to GTTAAC). A blunt-ended neomycin (*neo*) resistance gene cassette was released from pPGKneobpA (14) and ligated into a *Pml* I site located 273 bp downstream of the single base insertion. A 6.3 kb *Eco* RI/*Asp*718 fragment encompassing 5′ flanking sequences and exon 1 and 2 of the POMC gene was added using Bluescript KS (+/−) (Stratagene, CA) as a cloning vector. A herpes simplex virus thymidine kinase (tk) gene was placed further upstream in the opposite transcriptional orientation with respect to the POMC gene to permit negative selection. Both the *neo* and the tk cassettes (gifts of P. Soriano) are under the transcriptional control of the phosphoglycerate kinase-1 (PGK) promoter. The resulting replacement targeting vector, designated pPOMCX*4, contains a long 5′ arm of 8 kb and a short 3′ arm of 1.2 kb homologous with the endogenous POMC gene (Fig. 1B).

Electroporation and selection of ES cells

AB1 (provided by A.Bradley) and D3 (provided by T.Doetschman) embryonic stem (ES) cells were grown on γ-irradiated G418 resistant SNL 76/7 feeder cells (provided by A. Bradley) (4) on gelatinized plates. ES cells were cultured as previously described (15). Confluent 10 cm plates (approximately 2×10^7 cells) were trypsinized and dissagregated by repetitive pipetting and resuspended in 0.8 ml Hepes buffered saline (21 mM Hepes pH 7.0, 137 mM NaCl, 5 mM KCl, 0.7 mM Na_2HPO_4, 6mM glucose). Electroporations were performed in the presence of 25 μg of *Not* I digested linearized pPOMC*4 at 280 V and 330 μF at room temperature using a cell porator (Bethesda Research Laboratories, MD). After 10 min, ES cells were divided into 3 equal aliquots and plated on 10 cm plates containing feeder cells. Following 24 to 36 h the medium was changed and contained G418 (300 μg/ml; Gibco-BRL, MD). Two plates were also treated with gancyclovir (2 μM; Syntex, CA). Media were changed every 48 h and after 8 to 10 days double resistant colonies were collected under a dissecting microscope with a fine Pasteur pipette and expanded as described elsewhere (16).

Identification of homologous recombinant clones by PCR and Southern hybridization

From each electroporation, 150 double resistant ES cell colonies were grown until confluence on 24-well plates. DNA was isolated as described previously (17) and resuspended in 0.15 ml of a solution containing 10 mM Tris–HCl and 0.1 mM EDTA, pH 7.5. PCR conditions were established on cells that had previously been transfected with a mock targeting construct. For the first screening round, pools were obtained combining DNA from 10 individual colonies. A second screening round was performed using DNA samples from positive pools differently combined in pools of 6 individual colonies. Finally the colonies present in the positive pools were individually analyzed. Typically PCR was carried out in a 50 μl final volume with 1 μl of each sample of DNA and the following mixture: 50 mM KCl, 10 mM Tris–HCl pH 8.3, 1.5 mM $MgCl_2$, 0.001% (w/v) gelatin, 200 μM of each dNTP, 1 μM of each primer and 1.25 U of Amplitaq DNA Polymerase, (Perkin Elmer Cetus, CA). Temperature cycles was as follows: 94°C, 1 min; 60°C, 1 min; 72°C, 1 min. Forty amplification cycles were used for the first and second screening and 35 for the third round. The following oligonucleotides were used to identify homologous recombination events: Primer 1 (Fig.1C): PGKneobpA sense (5′GAGG-ATTGGGAAGACAATAGCA3′) and primer 2: 3′ end of the POMC genomic clone outside of the targeting vector antisense (5′GACATGTTCATCTCTATACATAC3′). PCR products were resolved in a 0.8% agarose gel and those showing a 1.2 kb band were treated as putative positives. 10 μg of genomic DNA from individual clones were digested with 20 U *Eco* RI at 37°C overnight, ethanol precipitated, run on a 0.8% agarose gel, electrotransferred to Nytran (Schleicher & Schuell, OH) and hybridized with a ^{32}P-labeled random primed (Boeringher-Manheinm, IN) probe using a 0.9 kb *Hind* III/*Eco* RI fragment located 3′ to the short arm of the targeting vector (probe A, Fig.1C). After developing the autoradiograms, the blots were stripped and rehybridized with a 0.8 kb *Pst* I/*Xba* I *neo* probe. Hybridizations were performed overnight at 42°C in the following conditions: 50% formamide, 6× SSC (1×SSC is 0.15 M NaCl, 0.015 M Na citrate, pH 7.2), 5×Denhardt's, 0.5% SDS, 25 mM sodium phosphate buffer pH 7.6, 100 mg/ml sheared and denatured salmon spearm DNA and 2.5 mM EDTA. Washes were performed at a final stringency of 0.1×SSC, 0.1% SDS at 65°C. To confirm the retention of the single base insertion, PCR was performed with each homologous recombinant clone using the same conditions as described above but with primers 3: POMC exon 3 sense (5′GGCAAGCGCTCC-TAC3′) and 4: PGK promoter antisense (5′CTAAA-GCGCATGCTCCAGAC3′). The 576 bp products were digested with *Hpa* I and resolved on a 1.5% agarose gel. Direct sequencing by the dideoxy chain termination method (Sequenase 2.0; USB, OH) of the PCR products was performed to assure that unwanted mutations had not occurred in the POMC coding sequences.

Analysis of POMC gene expression in homologous recombinant clones

Two homologous recombinant clones and wild type D3 cells were grown to confluence on 10 cm gelatinized plates in the presence of 1000 U/ml murine leukemia inhibitory factor (Esgro, Gibco–BRL, MD) and total RNA was prepared using the proteinase K/NP-40 lysis method described elsewhere (18). Similarly, total RNA was isolated from three pooled C57Bl/6 mouse pituitaries. 20 μg of total RNA from each sample were loaded into a formaldehyde gel (18), run at 3 V/cm for 4 h, electrotransferred to Nytran and hybridized with the *neo* probe as described above. After developing the autoradiogram, the blot was stripped and rehybridized with a ^{32}P-labeled random primed probe using a 0.5 kb *Nco* I/*Pml* I fragment present in the third exon of the mouse POMC gene. Hybridization and washing conditions were the same as described for Southern blots.

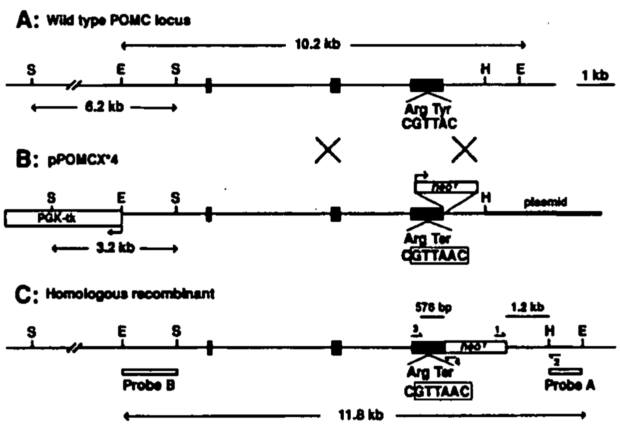


Figure 1. Schematic representation of a point mutation introduced in the POMC gene by homologous recombination. A: Restriction map of the wild type POMC locus in the region homologous to pPOMCX*4. Black boxes denote exons. B: Restriction map of the linearized targeting construct pPOMCX*4 after *Not* I digestion. A *neo* cassette was inserted after the natural polyadenylation signal of the POMC gene and a tk cassette was added at the 5′ end. The plasmid contains 8.1 kb of homology 5′ and 1.2 kb 3′ of the *neo* insertion. The sequence in the box (GTTAAC) denotes the *Hpa* I restriction site created after the insertion of a single A. Arrows indicate transcriptional orientation of the *neo* and tk promoters. C: Predicted structure of the mutated allele after homologous recombination. Half arrows indicate the approximate positions of primers 1 and 2 used for PCR amplification in colony screening and primers 3 and 4 used to confirm the retention of the point mutation. Probes A and B were used in Southern blots to identify homologous recombination events. Abbreviations: E, *Eco* RI; H, *Hind* III; B, *Bam* HI; S, *Sma* I.

RESULTS AND DISCUSSION

Targeting a single base insertion to the POMC gene by homologous recombination

We used a 10.2 kb *Eco* RI/*Eco* RI fragment isolated from a Balb/c mouse genomic DNA library (13) to target the POMC locus. This fragment contains the complete transcriptional unit of the POMC gene (Fig. 1A). With the ultimate aim of creating mice capable of producing normal ACTH and MSH peptides but unable to produce functional β-endorphin, we constructed the replacement targeting vector pPOMCX*4 (Fig. 1B) which carries two independent mutations. First, there is a single base insertion in the first codon of β-endorphin that converts the amino-terminal tyrosine into a premature translational termination codon (TAC to TAA). Because the coding region for β-endorphin is located 3′ with respect to ACTH and MSH sequences, the predicted carboxyl-terminally truncated preprohormone is likely to be processed to all possible POMC peptides with the exception of β-endorphin and its derivatives. Our strategy makes the assumption that the truncated prohormone will be correctly targeted to secretory granules, however the location of the signal(s) for intracellular trafficking within POMC are not known. The second mutation is the insertion of a *neo* expression cassette PGKneobpA (14) that is used to select transformed cells. *Neo* was placed downstream of the natural polyadenylation sequence of the POMC gene and 1.2 kb from the 3′ end of the construct. This positioning permitted the identification of potential homologous recombinants using PCR with a primer complementary to the sense strand of the *neo* cassette (primer 1, Fig. 1C) and another one complementary to the antisense strand of a region of the POMC locus not present in the targeting construct (primer 2). A herpes simplex virus thymidine kinase expression cassette (PGKTK) was added at the 5′ end of the long arm of the construct to select against random integration events. pPOMCX*4 was used to electroporate 2×10^7 AB1 or D3 ES cells and the transformants were selected in G418 and gancyclovir. The enrichment obtained with gancyclovir was 3.9 fold in AB1 cells and 1.6 fold in D3 cells (Table 1).

To detect homologous recombinant clones, we used PCR primers specific for a junction fragment at the 3′ end of the construct (Fig. 1C). In order to avoid the presence of feeder cells during the preparation of genomic DNA, ES cells were grown during the last passage in the presence of 1000 U/ml murine leukemia inhibitory factor. PCR positive clones were confirmed by Southern blot analysis of genomic DNA digested with *Eco* RI and hybridized to probe A, a flanking probe in the 3′ vicinity of the targeting vector (Fig. 1C). Ten of 11 positive clones identified initially by PCR showed the expected restriction pattern: a 10.2 kb band corresponding to the normal allele as well as the

Table 1. Introducing a point muation and a selectable marker in the POMC locus

ES Cells	No. of Cells Electroporated	No. of Colonies $G418^r$	$G418^r/GANC^r$	Enrichment	No. of Recombinants	Frequency of Recombination	Retention of the Point Mutation
AB1	2×10^7	3,400	880	3.9	5/140	1/28	4/4
D3	2×10^7	720	460	1.6	6/144	1/24	5/6
Total	4×10^7	4,120	1,340	2.8*	11/284†	1/26	9/10†

* Represents the average obtained between both cell lines. GANC, gancyclovir.
† Eleven clones were identified by a PCR strategy. Ten of these had the correct structure based on subsequent Southern blot analysis.

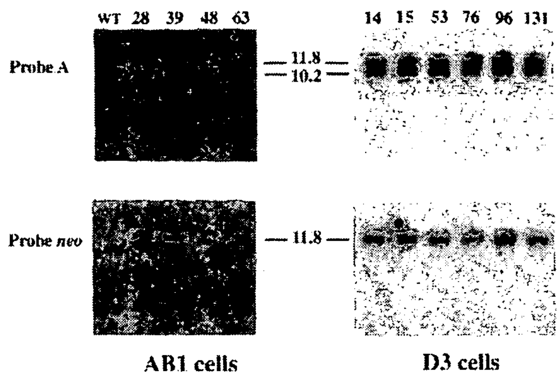


Figure 2. Southern blot analysis of genomic DNA from recombinant embryonic stem (ES) cell clones. Ten micrograms of DNA from wild type (WT) AB1 cells or ES cell clones identified as recombinants by an initial PCR screen were digested with *Eco* RI and sequentially hybridized with a ^{32}P-labeled random primed probe using a 0.9 kb *Hind* III/*Eco* RI fragment (probe A, Figure 1C) and with a ^{32}P-labeled random primed probe using a 0.8 kb fragment of *neo*. The 10.2 kb band corresponds to the wild type allele while the 11.8 kb band concides with the predicted homologous recombination event. Autoradiograms were exposed for 3 to 4 days.

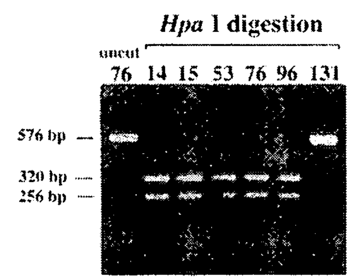


Figure 3. Analysis of the retention of the point mutation. DNA samples from D3 ES cell homologous recombinant clones were used as templates to PCR amplify a 576 bp band using primers 3 and 4 (shown in Figure 1C). These bands were further digested with *Hpa* I and separated on a 1.5% agarose gel. The size of the predicted bands are 320 bp and 256 bp. An undigested aliquot of the PCR product obtained with DNA from clone 76 was run on the left lane as a control.

predicted 11.8 kb band created following homologous recombination (Fig. 2). When the same blot was stripped and rehybridized with a *neo* probe, only the 11.8 kb band was observed suggesting that none of the targeted clones carried an additional integrated insertion of the construct or contained any gross structural rearrangement of the locus (Fig. 2). Homologous recombination at the 5′ extreme of the construct was confirmed also by Southern blot analysis of genomic DNA digested with *Bam* HI and *Sma* I and hybridized to probe B (data not shown). Table 1 summarizes these experiments and shows a relatively high frequency of recombination at the POMC locus, 1/28 and 1/24 for AB1 and D3 cells respectively.

Retention of the single base insertion in the homologous recombinant clones

Although the mechanisms of homologous recombination in mammalian cells are not yet completely understood, some evidence suggests that the process is similar to that found in bacterial and fungal systems: double reciprocal crossover events followed by the resolution of Holliday structures or by gene conversion (19). A typical replacement vector contains regions of homology with the endogenous locus at both sides of the mutated sequence and the vector is thought to pair via shared homology with its chromosomal counterpart.

Because the targeting vector pPOMCX*4 carries two different mutations, it was necessary to confirm the existence of the single base insertion responsible for knocking out β-endorphin production in those clones selected for homologous recombination by G418 and gancyclovir resistance. Using genomic DNA from those clones and primers 3 and 4 (Fig. 1C) we PCR amplified a 576 bp product that contains the first codon of β-endorphin and the newly created *Hpa* I restriction endonuclease site. *Hpa* I digestion should generate two fragments of 320 bp and 256 bp. All the recombinant clones in AB1 ES cells analyzed (data not shown) as well as five out of six D3 recombinant clones carried the single base mutation. Figure 3 shows the restriction pattern of the PCR products amplified with primers 3 and 4 after digestion with *Hpa* I. Clones 14, 15, 53, 76 and 96 had the predicted 320 and 256 bp bands while the uncut band of clone 131 suggests that the *Hpa* I site was absent. Direct sequencing of the 576 bp PCR products confirmed the predicted mutation in all the clones that were cut with *Hpa* I. However, clone 131 had the same sequence as the wild type POMC gene (data not shown). The presence of a recombinant that carries only one of the mutations could be due to either a cross-over event on the long arm between both mutations or an initial cross-over event upstream of the point mutation, with restoration of the wild type sequence by branch migration. The high retention of the point mutation in 9 out of 10 recombinant events may be due to the relatively short distance between the insertional point mutation and the PGK*neo* cassette (273 bp). Further experiments are necessary to distinguish between these possibilities.

A limited number of strategies have been reported to introduce small mutations in the mammalian germ line. They include the 'it and run' (7) similar to the 'in-out' (8) method and coelectroporation of a targeting replacement vector together with a selectable marker in a different plasmid (9). The first strategy is essentially a double recombination. This procedure is very complicated because the positive clones obtained after the initial homologous recombination between an insertional targeting vector and the endogenous locus must be reselected for intrachromosomal recombination between the endogenous

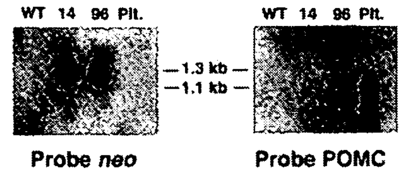


Figure 4. Northern blot analysis of POMC and PGK promoter transcriptional activities in embryonic stem cells. Total RNA from wild type D3 ES cells (WT), from homologous recombinant D3 clones 14 and 96 and from normal mouse pituitaries (Pit.) was obtained and 20 μg were loaded per lane. The Northern blot was first hybridized with a ^{32}P-labeled random primed probe using a 0.8 kb fragment of *neo*. The autoradiogram was exposed for 4 d. After stripping the blot was rehybridized with a ^{32}P-labeled random primed probe using a 0.5 kb *Nco* I/*Pml* I fragment of the third exon of the mouse POMC gene and the autoradiogram was exposed for 5 h.

sequences and the duplicated sequences provided by the vector. Coelectroporation, on the other hand, has the disadvantage of limiting the primary screening for targeted events in a large number of clones to the inefficient approach of direct sequencing of PCR products. Although the introduction of a restriction site allows screening of recombinants by Southern hybridization or endonuclease digestions of PCR products that encompass the mutation, most point mutations will not generate or disrupt a restriction site.

The methods described above, although using more complex or time consuming routines, do avoid the presence of a selectable marker with its promoter/enhancer in the genome of the mutated animal. Potentially, the promoter of the selection cassette could positively or negatively regulate expression of the targeted locus thereby complicating an interpretation of the phenotype resulting from a subtle mutation in the encoded protein sequences. In addition, the PGK promoter was reported to have limited bidirectional activity in transfected cell lines (20). For these reasons, we tested the capability of the PGK promoter to influence the expression of the POMC gene in ES cells targeted with pPOMCX*4. Total RNA was prepared from wild type and from two different homologous recombinants of D3 ES cells and from mouse pituitaries and used for Northern blots of *neo* and POMC transcripts. Figure 4 shows that a 1.3 kb band predicted for *neo* mRNA is only present in recombinant ES clones suggesting that the PGK promoter is active in these cells. A random primed probe using double-stranded POMC DNA as a template was used to detect potential POMC transcripts made in either orientation. After stripping and rehybridizing the Northern blot, only the lane containing RNA from mouse pituitaries showed the predicted 1.1 kb band corresponding to mouse POMC mRNA. Neither wild type nor the recombinant ES cells showed detectable levels of POMC gene expression even after 48 h of film exposure, suggesting that the active PGK promoter was unable to inappropriately activate the normally silent POMC gene in ES cells.

It is possible, therefore, that the retained PGK-*neo* cassette will have no effect on POMC gene expression in differentiated tissues of mice produced from the targeted embryonic stem cells. In at least one case where expression was assayed, a retained polyoma promoter/enhancer-*neo* selection cassette was actually silent *in vivo* and the disrupted insulin-like growth factor-2 allele was transcribed essentially in the normal developmentally regulated and tissue-specific pattern (21). On the other hand ectopic expression of a selectively mutated allele would not necessarily confound an interpretation of the phenotype. Several examples of transgenic mice with ectopic expression of a neuropeptide and no overt phenotype have been reported (22–25).

Our results demonstrate that it is possible to introduce precisely a point mutation into the mouse genome using a replacement type vector with a retained selectable marker. The absolute frequency of correctly targeted clones was at least 10-fold higher than reported with other more complicated strategies (7–9). POMC, like other neuropeptide precursors, has multiple proteolytic processing sites that are typically paired basic amino acid residues. Mutation of the Lys-Arg processing site flanking β-endorphin could therefore be an alternative approach to producing mice with a selective loss of the opiate peptide if the carboxyl-terminal truncated POMC prohormone encoded by the mutation described in this report is not properly targeted or processed. A similar mutational strategy could be used to target individual or multiple processing sites and selectively alter the resulting peptide profile generated from prohormones to deduce the physiological functions of each peptide in mice. In the example reported, the point mutation was introduced into the terminal exon and the *neo* cassette was inserted into the 3′ flanking region of the gene. Alternatively, it should be possible to insert the *neo* cassette into an intron to accomodate a small mutation in a central exon and maintain a short distance between the two mutations. The relative simplicity and efficiency of this strategy will likely make it applicable to a larger number of genes than the previously described methods.

ACKNOWLEDGMENTS

We thank A.Bradley, T.Doetschman, P.Soriano and M.Uhler for providing the essential cell lines, selectable marker cassettes, POMC genomic clone and for valuable advice. This work was supported by an NIH grant DK 40457 to M.J.L. This work was partially supported by Fondo de Investigación Sanitaria de la Seguridad Social (BAE 91/5879) to M.A.J. and Fundación Autorchas to M.R.

REFERENCES

1. Snouwaert, J. N., K. K. Brigman, A. M. Latour, N. N. Malouf, R. C. Boucher, O. Smithies, and B. H. Koller. (1992) *Science* 257, 1083–1088.
2. Tybulewicz, V. L. J., M. L. Tremblay, M. E. LaMarca, R. Willemsen, B. K. Stubblefield, S. Winfield, B. Zablocka, E. Sidransky, B. M. Martin, S. P. Huang, K. A. Mintzer, H. Westphal, R. C. Mulligan, and E. I. Ginns. (1992) *Nature* 357, 407–410.
3. Thomas, K. R., and M. R. Capecchi. (1990) *Nature* 346, 847–850.
4. McMahon, A. P., and A. Bradley. (1990) *Cell* 62, 1073–1085.
5. Lufkin, T., A. Dierich, M. LeMeur, M. Mark, and P. Chambon. (1991) *Cell* 66, 1105–1119.
6. Li, E., T. H. Bestor, and R. Jaenisch. (1992) *Cell* 69, 915–926.
7. Hasty, P., R. Ramírez-Solís, R. Krumlauf, and A. Bradley. (1991) *Nature* 350, 243–246.
8. Valancius V., and O. Smithies. (1991) *Mol. Cell. Biol.* 11, 1402–1408.
9. Davis, A. C., M. Wims, and A. Bradley. (1992) *Mol. Cell. Biol.* 12, 2769–2776.
10. Zimmer, A., and P. Gruss. (1989) *Nature* 338, 150–153.
11. De Wied, D., and J. Jolles. (1982) *Physiol. Rev.* 62, 976–1059.
12. Smith, I. A., and J. W. Funder. (1988) *Endocrine Rev.* 9, 135–158.
13. Uhler, M., E. Herbert, P. D'Eustachio, and F. D. Ruddle. (1983) *J. Biol. Chem.* 258, 9444–9453.
14. Soriano, P., C. Montgomery, R. Geske, and A. Bradley. (1991) *Cell* 64, 693–702.
15. Robertson, E. J. (1987) *In* E. J. Roberston (ed.), Teratocarcinomas and embryonic stem cells: a practical approach. IRL Press, Oxford, pp. 71–112..
16. Ure, J. M., S. Fiering, and A. G. Smith. (1992) *Trends Genetics* 8, 6.
17. Laird, P. W., A. Zijderveld, K. Linders, M. A. Rudnicki, R. Jaenisch, and A. Berns. (1991) *Nucl. Acid Res.* 19, 4293.
18. Sambrook, J., E. F. Fritsch, and T. Maniatis. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. Bollag, R. J., A. S. Waldman, and R. M. Liskay. (1989) *Annu. Rev. Genet.* 23, 199–225.
20. Johnson, P., and T. Friedman. (1990) *Gene* 88, 207–213.
21. DiChiara, T. M., E. J. Robertson, and A. Efstratiadis. (1991) *Cell* 64, 849–859.
22. Low, M. J., R. E. Hammer, R. H. Goodman, J. F. Habener, R. D. Palmiter, and R. L. Brinster. (1985) *Cell* 41, 211–219.
23. Günsburg, W. H., B. Salmons, B. Zimmerman, M. Müller, V. Erfle, and G. Brem. (1991) *Mol. Endocrinol.* 5, 123–133.
24. Russo, A. F., E. B. Crenshaw III, S. A. Lira, D. M. Simmons, L.W. Swanson, and M. G. Rosenfeld. (1988) *Neuron* 1, 311–320.
25. Rubinstein, M., B. Liu, R. H. Goodman, and M. J. Low. (1992) *Mol. Cell. Neurosci.* 3, 152–161.

Exhibit 3

MOLECULAR AND CELLULAR BIOLOGY, Mar. 1991, p. 1402–1408 Vol. 11, No. 3
0270-7306/91/031402-07$02.00/0

Testing an "In-Out" Targeting Procedure for Making Subtle Genomic Modifications in Mouse Embryonic Stem Cells

VICKY VALANCIUS AND OLIVER SMITHIES*

Department of Pathology, The School of Medicine, University of North Carolina, Chapel Hill, North Carolina 27599-7525

Received 10 September 1990/Accepted 18 December 1990

We have introduced a 4-bp insertion into the hypoxanthine phosphoribosyltransferase (*HPRT*) gene of a mouse embryonic stem (ES) cell line by using an "in-out" targeting procedure. During the in step, a homologous integration reaction, we targeted a correcting plasmid to a partially deleted *hprt*$^-$ locus by using an integrating vector that carried a 4-bp insertion in the region of DNA homologous to the target locus. *HPRT*$^+$ recombinants were isolated by direct selection in hypoxanthine-aminopterin-thymidine (HAT) medium. The HAT^r cell lines were then grown in medium containing 6-thioguanine (6-TG) to select for *hprt*$^-$ revertants resulting from the excision of the integrated vector sequences. The revertants were examined by Southern blot hybridization to determine the accuracy of this out reaction and the frequency of retaining the 4-bp modification in the genome. Of the 6-TG^r colonies examined, 88% had accurately excised the integrated vector sequences; 19 of 20 accurate revertants retained the 4-bp insertion in the resulting *hprt*$^-$ gene. We suggest a scheme for making the in-out targeting procedure generally useful to modify the mammalian genome.

Gene targeting, the method of using homologous recombination to modify the mammalian genome, can be used to introduce specific changes into cultured cells. By targeting the gene of interest in embryonic stem (ES) cells, these changes can be introduced into the germ lines of laboratory animals to study the effects of the modifications on whole organisms. During the gene targeting procedure, cloned DNA that is homologous to the target locus but has been modified in vitro to carry the intended change is introduced into tissue culture cells. The treated cells are then screened for accurate targeting to find those that have been properly modified.

Several schemes have been devised to aid in the isolation of the recombinants. A positively selectable helper gene is often included in the targeting vector to permit selection for cells that have taken up and expressed the introduced DNA. Typically, this selection results in a 1,000-fold enrichment for targeted cell lines (7, 10, 11, 20, 24). Addition of a second, negatively selectable helper gene allows for selection against random insertion events (15). The positive-negative selection scheme achieves an additional 10- to 1,000-fold enrichment for the homologous recombinants over the single positive selection procedure (4, 15). In both the single positive and the dual positive-negative selection systems, the positive selection element is introduced into the target locus in such a way as to interrupt or knock out the targeted region.

To create subtle modifications of the mammalian genome for fine-structure analyses or to mimic human genetic diseases resulting from small specific mutations, it is desirable to develop a targeting procedure that allows for easy identification of the recombinant cell lines (such as with the aid of selectable sequences) yet does not leave extraneous sequences in either the target or any other region of the genome in the final product. Recently, Steeg et al. (21) reported the introduction of a single base pair change into a mammalian genome without accompanying alterations. They targeted the RNA polymerase II gene in a mouse ES cell line in such a way that homologous recombinants, different from the parental line by only one or two base pairs, were directly selectable with α-amanitin. Zimmer and Gruss (27) described the insertion of 20 base pairs into the Hox 1.1 locus in mouse ES cells without using any selection. The DNA was introduced by nuclear microinjection, and the polymerase chain reaction was used to detect recombinants. To date, there have been no other reports of successful gene targeting using this procedure.

Another procedure for making subtle changes in loci which are not directly selectable was described for the yeast system by Scherer and Davis (17). They targeted the *HIS3* locus with an insertional plasmid carrying a mutated *his3* gene as well as the directly selectable *URA3* gene. Targeted recombinants, isolated by the URA3$^+$ phenotype, were expanded without selection then screened for the ura3$^-$ phenotype. They were able to isolate colonies that had lost all target vector sequences but now had only the mutated *his3* gene initially introduced into the genome by the targeted insertion event. As other yeast targeting techniques have been successfully transferred to the mammalian system, it should be possible to use this type of two-step selection scheme to make subtle changes in nonselectable loci in mammalian cells.

We have therefore tested the feasibility of using a two-step selection procedure in mouse ES cells to introduce a small change into a target locus. To this end, we targeted the mutant hypoxanthine phosphoribosyltransferase (*hprt*$^-$) gene in the cell line E-14TG2a (8) with an insertional plasmid capable of correcting the deletion mutation at this locus. The plasmid carried a 4-bp insertion in the region of DNA homologous to the target locus. *HPRT*$^+$ targeted cell lines were isolated by direct selection with hypoxanthine-aminopterin-thymidine (HAT) medium, expanded, and then selected with 6-thioguanine (6-TG) to recover *hprt*$^-$ revertants. The revertant colonies were examined by genomic Southern blot hybridization to confirm that the integrated vector DNA had been excised from the genome and to determine the frequency at which the intended 4-bp modifi-

* Corresponding author.

cation was obtained. Our results show that a subtle modification of the mouse *HPRT* gene can be achieved by using the two-step selection procedure. This "in-out" targeting procedure can be adapted to modify nonselectable loci with the aid of the *HPRT* minigene constructed by Reid et al. (16).

MATERIALS AND METHODS

Cell culture. The mouse ES cell line E-14TG2a was isolated as described previously (8, 25). Cells were grown in Dulbecco's modified Eagle's medium (GIBCO) supplemented with 15% heat-inactivated fetal calf serum (Flow) and 10 μM 2-mercaptoethanol (Sigma). The pluripotential nature of the ES cells was retained by supplementing each liter of growth medium with 10^6 U of recombinant human leukemia inhibitory factor, graciously supplied by N. Gough (Walter and Eliza Hall Institute, Melbourne, Victoria, Australia). Because feeder layers were not used, all culture dishes were coated with 0.1% sterile gelatin to ensure cell adhesion. HAT medium was standard culture medium supplemented with 120 μM hypoxanthine, 0.4 μM aminopterin, and 20 μM thymidine. 6-TG selection was carried out in standard medium containing 10 μM 6-TG. Cultures were incubated at 37°C in an atmosphere of 5% CO_2. They were checked periodically for mycoplasma contamination.

Vectors. Plasmid pNMR133 has already been described (6). It contains 5 kb of DNA identical to the exon 3 target region of the mouse *HPRT* gene, except for a 4-bp insertion that destroys a unique *Hin*dIII site and consequently generates a new *Nhe*I site in intron 2. It also carries the human *HPRT* promoter and exon 1 sequences (which have been shown to function in mouse cells) and the mouse exon 2 region.

Plasmid pNMR133D200 was derived from pNMR133 by removing a 200-bp *Bgl*II fragment from intron 2.

DNA preparation. Targeting vector DNAs were prepared by standard methods, omitting the CsCl purification, which we found unnecessary (unpublished results). All targeting DNAs were linearized by restriction enzyme digestion, using the manufacturers' recommended conditions, prior to electroporations. Digested DNAs were ethanol precipitated and resuspended in sterile TE buffer (0.05 M Tris, 0.001 M EDTA).

DNA transfers and selections. The vectors were introduced into the ES cells by electroporation (2). The cells were grown in 100-mm culture dishes (as described above) to a density of 1×10^7 to 2×10^7 cells per dish in nonselective medium. Cultures were trypsinized, centrifuged, and then resuspended in nonselective medium to a density of 4×10^7 to 10×10^7 cells per ml. A 0.5-ml sample of the cell suspension was added to each microfuge tube, and prepared DNA was then added to a final concentration of 5 nM. The cell-DNA mixtures were incubated on ice for 20 min, loaded into an electroporation chamber precooled on ice (length, 5 mm; cross section, 100 mm^2), and exposed to a 1-s electrical pulse from a 250-μF capacitor charged to 300 V. Cells were immediately removed from the chamber and plated into five 100-mm culture dishes. The plates had been prepared by gelatinization and contained 7 ml of nonselective medium. The cells were allowed to recover overnight. The next day, the number of colonies in each dish was determined by counting, and HAT selection was then applied.

Cultures to be assayed for the loss of *HPRT* function by selection in 6-TG were maintained under HAT selection for at least 1 month prior to the start of the assay in order to kill any accumulated *hprt*$^-$ cells. These cultures were trypsinized, counted, and then replated at a density of 0.5×10^7 to 1×10^7 cells per plate in nonselective medium. They were grown without selection for 3 or 4 days to allow spontaneous revertants time to purge residual *HPRT* transcripts or protein. Selection was then started by applying 6-TG medium.

All selections were maintained for 16 days, with feeding as necessary. Targeting and reversion frequencies were determined by counting the number of resistant colonies obtained for each experiment. Individual colonies were picked by using cloning rings into 24-well (1 ml per well) dishes and maintained under selection. Cultures were transferred to 60-mm culture dishes and then either harvested for genomic DNA preparation or transferred to 100-mm dishes for further expansion.

Genomic DNA preparation and characterization. DNA was prepared from expanded clones by using conventional procedures. Restriction enzyme digestions were done according to manufacturers' specifications, incubating overnight. After electrophoresis on 0.8% agarose gels, Southern blotting was done by standard techniques.

Probes. Two probes were used, a 250-bp *Rsa*I fragment from intron 3 and a 300-bp *Hin*dIII-*Xho*I fragment from the human cDNA which includes exons 3 to 6 but is specific for the mouse exon 3 element (6). Both probes hybridize to sequences present in the endogenous locus as well as on the targeting vectors. For each blot, 25 to 50 ng of purified fragment was radiolabeled with ^{32}P-dCTP by the random-primed oligonucleotide method, using a Boehringer Mannheim kit. Four-hour prehybridizations and overnight hybridizations were done in 50% formamide solutions at 42°C. Blots were washed to a stringency of 1× SSC (0.15 M NaCl plus 0.015 M sodium citrate) at 68°C. Washed blots were exposed to preflashed XAR-5 film at −70°C.

RESULTS

In step: the integration event. The first step in the two-step targeting procedure is a homologous integration event that incorporates vector DNA carrying the desired modification into the genome. We used the method of Doetschman et al. (6) to introduce into mouse ES cells an integrating targeting vector that carries a 4-bp insertion in the second intron of the *HPRT* gene. Either plasmid pNMR133 or plasmid pNMR133D200 (which has a 200-bp gap in the region homologous to the target locus) was electroporated into the male mouse-derived ES cell line E-14TG2a. This cell line, isolated by Hooper et al. (8) as a spontaneous mutation in culture, carries a nonreverting deletion of the promoter and first two exons of the nine-exon, 33-kbp, X-linked *HPRT* gene (25), rendering it phenotypically *hprt*$^-$. Both targeting vectors contain approximately 5 kb of DNA identical in sequence to the exon 3 target region of the *hprt*$^-$ gene except for the intended modification: a 4-bp insertion in intron 2 that destroys a unique *Hin*dIII site. They also carry the human *HPRT* promoter and exon 1 sequences and the mouse exon 2 region.

The homologous integration event generates a duplication of the 5-kb target region separated by the remainder of the vector sequences (Fig. 1). The duplicated regions are identical with the exception of the 4-bp insertion, identified by a missing *Hin*dIII site, that is located on the downstream repeat. This event restores the promoter and first two exons deleted from the locus, generating *HPRT*$^+$ targeted recombinants that can be directly selected with HAT-containing medium.

We isolated three independent *HPRT*$^+$ cell lines by selec-

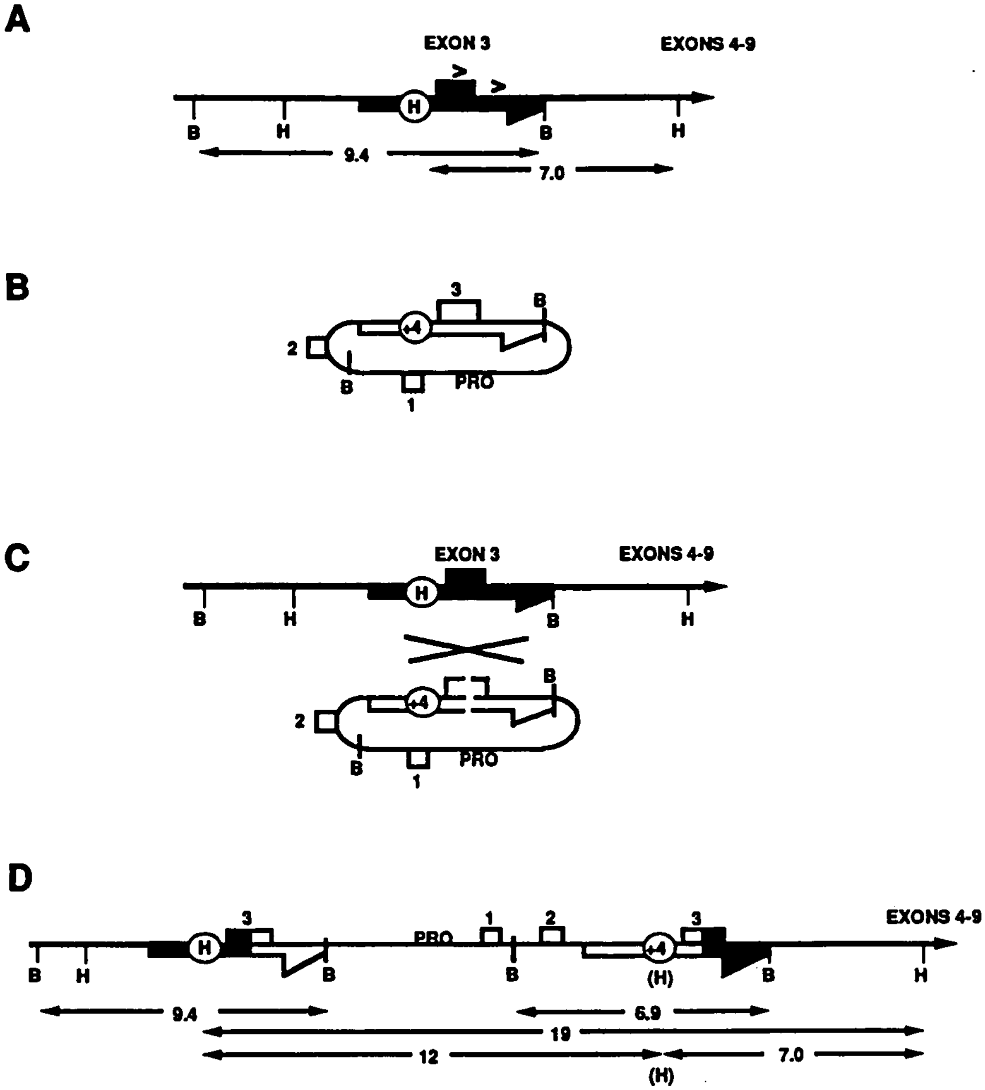


FIG. 1. The integration reaction. (A) The *hprt*⁻ locus found in cell line E-14TG2a. The 5-kb target region is denoted by the thick black arrow. >, Regions used as probes. These sequences are also present in the targeting vectors. (H), The *Hin*dIII site that has been destroyed in the targeting vectors. (B) The pNMR133 targeting vector (12 kb). +4 denotes the 4bp insertion carried by this plasmid; it destroys the *Hin*dIII ((H)) site. Plasmid pNMR133D200 is identical to this plasmid except that a 200-bp *Bgl*II fragment has been removed from the region of homology. The open arrow denotes the region of DNA homologous to the target locus. PRO, Promoter; 1, 2, and 3, exons. (C) The homologous integration reaction with pNMR133 linearized at the *Xho*I site in the middle of the *HPRT* sequences. (D) The resulting *HPRT*⁺ recombinant locus. The 5-kb duplicated regions (the large half-black, half-open arrows) are composed of both target locus and vector-derived sequences. B, *Bam*HI; H, *Hin*dIII.

tion in HAT medium at an average frequency of 2.8×10^{-6} per electroporated cell and then confirmed that these cell lines were targeted by genomic Southern blot hybridization. The blots were probed either with a 250-bp *Rsa*I fragment from intron 3 or with a 300-bp *Hin*dIII-*Xho*I fragment from the human cDNA that specifically hybridizes to the mouse exon 3. Both probes hybridize to sequences found in the genome as well as on the targeting vectors (Fig. 1A). All of the cell lines examined contained the expected recombinant locus depicted in Fig. 1D, indicating that a single copy of the targeting vectors had integrated into the E-14TG2a *hprt*⁻ gene.

Cell lines A and C hybridized to the single 19-kb *Hin*dIII fragment expected for a simple insertion of the 12-kb vector into the 7-kb endogenous fragment. Cell line D hybridized to two *Hin*dIII fragments, the endogenous 7-kb and the vector 12-kb fragments. This cell line, generated with plasmid pNMR133D200, has lost the 4-bp insertion as a consequence of the integration event (see reference 6), so that revertants obtained from this line could not be properly modified. However, it was used in the excision experiments (see below) since it could still generate useful information about the frequency and accuracy of the excision reaction. *Bam*HI digestion of all recombinants revealed the expected 9.4-kb

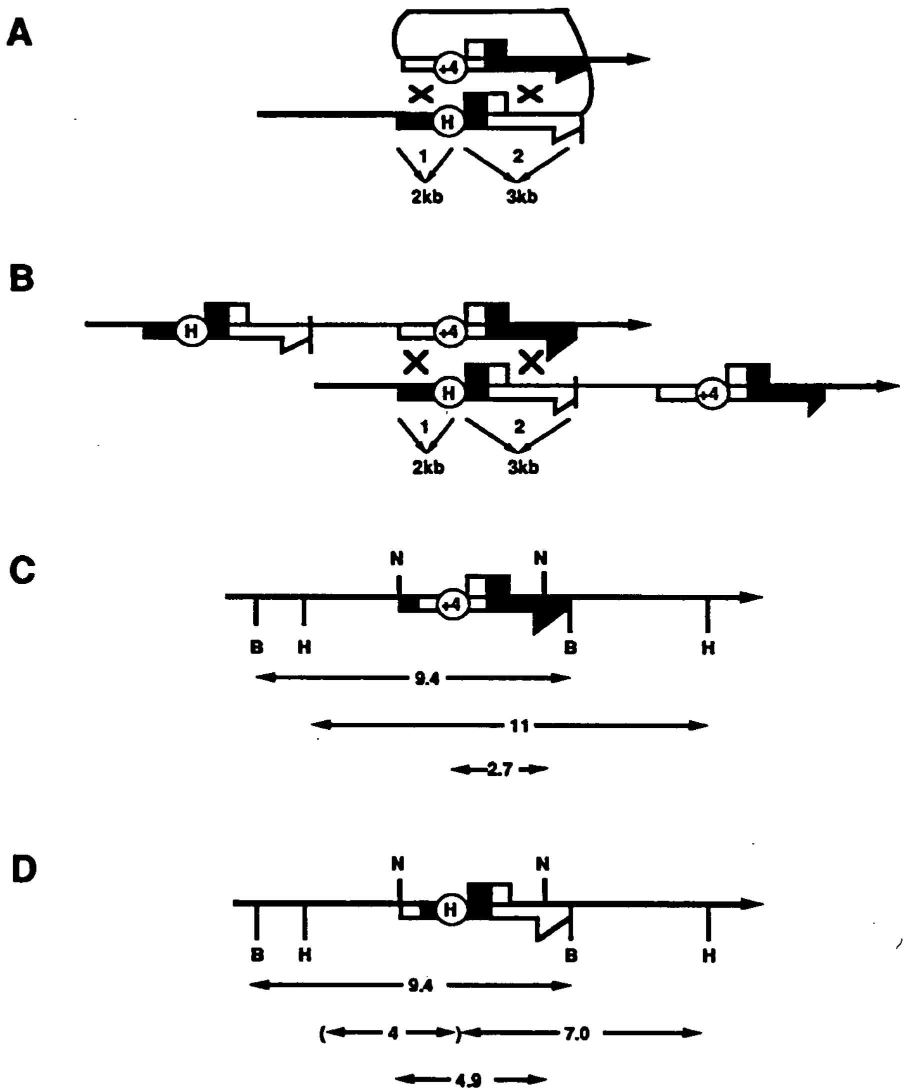


FIG. 2. The excision reaction. (A) Homologous excision via intrachromosomal recombination between the 5-kb duplications. Crossover in region 1 (2 kb) retains the 4-bp insertion in the excision product; crossover in region 2 (3 kb) removes the 4-bp insertion. (B) Homologous excision via unequal sister chromatid exchange. One of the resulting chromosomes (the excision product) contains a single copy of the duplicated region. Regions 1 and 2 are as in panel A. (C) The desired product: the *hprt*$^-$ locus now containing the 4-bp insertion. (D) The other possible *hprt*$^-$ locus: the E-14TG2a starting locus (Fig. 1A). The 4-kb fragment in parentheses does not hybridize to our probes; it is included in this figure to clarify the origin of the 11-kb band in panel C. B, *Bam*HI; H, *Hin*dIII; N, *Nhe*I.

endogenous band and the 6.9-kb vector-derived band. No extraneous bands could be detected, confirming that all of the recombinants carried single-site, single-copy insertions of the vector DNAs (see Fig. 3). In addition to these three lines, one more cell line, B, generated previously in the laboratory (6) was used in the excision studies (see below). This line carries the same recombinant locus found in lines A and C.

Out step: the excision event. The second step in the two-step targeting procedure is a spontaneous event that excises from the genome the vector sequences that integrated in the first step. A homologous recombination event between the regions duplicated during the in reaction can occur by either intrachromatid recombination (Fig. 2A) or unequal sister chromatid exchange (Fig. 2B). A crossover event in the 2-kb region 1 will leave the 4-bp insertion in the genome; crossing over in region 2, which is 3 kb, will excise the 4-bp modification along with the vector sequences (Fig. 2A) or move it to the (*HPRT*$^+$) triplicated chromosome (Fig. 2B). Either way, the excision event removes the vector-derived promoter and first two exons, causing a reversion to the *hprt*$^-$ phenotype. Such revertants can be selected with the nucleoside analog 6-TG.

The four HAT^r cell lines described above were used to study the excision (out) reaction. They all carry essentially the same *HPRT* locus: a duplication of 5 kb separated by 7 kb of plasmid-derived unique sequence (Fig. 1D). ES cell line D-3 (5), which carries the wild-type *HPRT* locus, was

TABLE 1. Frequency of the out reaction

Cell line	Cells plated (10^7)	Colonies 24 h post (10^6)[a]	6-TG^r colonies[b]	Reversion frequency[c]
A	2.90	3.2	23	7.9×10^{-7}
B	2.2	3.2	13	5.9×10^{-7}
C	4.2	6.4	14	3.3×10^{-7}
D	3.8	4.0	56	14.7×10^{-7}
Total	13.1		106	8.1×10^{-7}
D-3 (control)	5.6	4.8	0	$<1.8 \times 10^{-8}$

[a] Number of colonies counted the day after replating, reflecting a 10 to 20% plating efficiency (see text).
[b] Number of colonies counted after 2 weeks of 6-TG selection.
[c] Number of 6-TG^r colonies obtained per plated cell.

used as a control in these experiments to determine the spontaneous rate of mutation from *HPRT*$^+$ to *hprt*$^-$ at the normal locus. The experiments were performed as described in Materials and Methods. The day after replating, the number of colonies observable in each dish was determined by counting. Typically, ES cells form one colony for every 5 to 10 cells plated (Table 1). This is due to their propensity to form aggregates, not to a high death rate. That is to say, each colony found the day after replating is composed of 5 to 10 individual cells. Although this aggregation may interfere with our 6-TG selections as a result of metabolic cross-feeding (9), it cannot be avoided (see Discussion).

The number of 6-TG^r colonies obtained for each line examined and calculated reversion frequencies are listed in Table 1. As shown, all four HAT^r lines initially generated by gene targeting reverted to the *hprt*$^-$ phenotype at similar frequencies, averaging 8×10^{-7} 6-TG^r colonies isolated for every HAT^r cell plated. Control cell line D-3, which carries the wild-type *HPRT* locus, failed to produce any 6-TG^r colonies from 5.6×10^7 cells plated. Thus, the spontaneous mutation frequency at the *HPRT* locus, for cells preselected with HAT, is less than 1.8×10^{-8}. This result is consistent with the rate of 1.5×10^{-8} per cell generation reported for the locus by Caskey and Kruh (3).

Several of the individual 6-TG^r colonies were analyzed further by genomic Southern blot hybridization, using the *HPRT*-specific probes described above. The number of colonies examined from each line and a summary of the results obtained from the Southern blot hybridizations are presented in Table 2. Representative Southern blots are presented in Fig. 3.

Of a total of 26 6-TG^r colonies examined, 23 (88%) had executed the out reaction and accurately excised the integrated vector sequences from the genome, as determined by the genomic Southern blots. They all revealed a single 9.4-kb *Bam*HI band upon hybridization, the size predicted for a simple homologous excision event. This is the same *Bam*HI fragment found in the parental E-14TG2a *hprt*$^-$ locus. *Hin*dIII digestion of the revertant DNAs is expected to reveal one of two bands upon hybridization: either an 11-kb fragment, if the crossover occurs in region 1 (Fig. 2A and B) and the 4-bp insertion introduced by the in event is retained, or a 7-kb fragment, if the crossover occurs in region 2 and the modification is removed from the *hprt*$^-$ genome.

TABLE 2. Accuracy of the out reaction

Cell line	Colonies examined	Accurate revertants[a]	Revertants with 4-bp insert[b]
A	3	2	2/2
B	9	7	6/7
C	11	11	11/11
D	3	3	NA
Total	26	23	19/20

[a] Number of colonies containing the expected *hprt*$^-$ locus.
[b] Number of *hprt*$^-$ colonies that retain the 4-bp insertion presented as a fraction of the number of accurate revertants obtained.
NA, Not applicable; this cell line does not carry the 4-bp insertion.

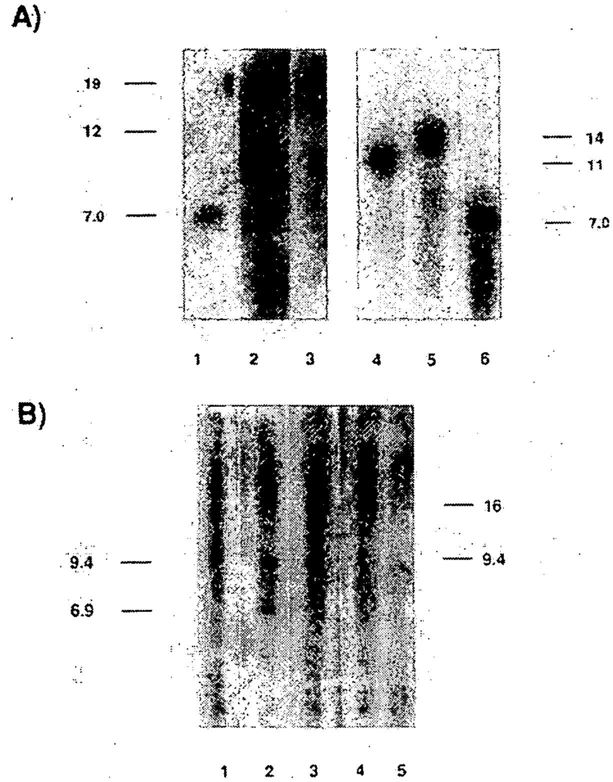


FIG. 3. Representative Southern blots. (A) *Hin*dIII-digested genomic DNAs hybridized to the exon 3 probe. Lanes: 1, starting cell line E-14TG2a (*hprt*$^-$); 2, HAT^r cell line without the 4-bp insertion (*HPRT*$^+$); 3, HAT^r cell line with the 4-bp insertion (*HPRT*$^+$); 4, correctly modified *hprt*$^-$ gene (with the 4-bp insertion); 5, unusual 6-TG^r revertant (see text); 6, accurate *hprt*$^-$ revertant that has lost the 4-bp insertion. (B) *Bam*HI-digested genomic DNAs hybridized to the 250-bp intron 3 probe. Lanes 1 and 2 are as in panel A, and lanes 3 to 5 are as lanes 4 to 6 in panel A.

Of the 23 out revertants examined, 20 were derived from *HPRT*$^+$ cell lines that carried the 4-bp insertion initially introduced by the targeted integration event; 19 of these colonies contain the single 11-kb *Hin*dIII fragment, indicative of the accurate excision event which retains the 4-bp insertion. Thus, these 19 colonies have been correctly modified by the in-out targeting procedure. Only 1 of these 20 revertant colonies lost the 4-bp modification, as determined by the presence of a 7-kb *Hin*dIII fragment. Therefore, 95% of the accurate revertants which could have retained the 4-bp insertion did so. The three remaining colonies which were found to have been generated by the out reaction were derived from the HAT^r cell line D. As this cell line does not carry the 4-bp modification, the revertants revealed only the 7-kb *Hin*dIII band upon hybridization.

To confirm that the 11-kb *Hin*dIII band characterizing our accurately modified $hprt^-$ revertants retained the 4-bp insertion initially introduced by the targeting vector, we digested two of the genomic DNAs with *Nhe*I. The 4-bp insertion introduced to destroy the *Hin*dIII site in the original targeting vector generates a unique *Nhe*I site. In the case of the revertants that have retained the 4-bp insertion (11-kb *Hin*dIII), *Nhe*I digestion will generate a 2.7-kb band which hybridizes to our probes. In the case of the revertants which have lost the insertion (7-kb *Hin*dIII), a 4.9-kb band will result (Fig. 2). The accurately modified revertant does contain a 2.7-kb *Nhe*I fragment which hybridizes to our probe, confirming the presence of the 4-bp insertion, and the revertant which has lost the 4-bp insertion reveals a 4.9-kb fragment upon hybridization (data not shown).

The other three colonies examined were found to contain aberrant $hprt^-$ loci that did not arise by the predicted homologous excision reaction. They contained a single 14-kb *Hin*dIII fragment and a 16-kb *Bam*HI fragment that hybridized to our probes. These fragments failed to hybridize to a plasmid-specific probe (data not shown), indicating that the target vector sequences have been excised from the genome. Since the bands are not the expected sizes, these colonies were probably generated by an alternate excision reaction. Because they account for only 12% of the 6-TG^r colonies obtained, we did not examine them further.

DISCUSSION

We have succeeded in modifying the genome of a mouse ES cell line, introducing a 4-bp insertion, by using a two-step in-out targeting procedure. The average frequency of the in reaction was found to be 2.8×10^{-6}, in agreement with other reports for gene targeting in murine systems (6, 7, 10, 20, 21, 23, 24). The frequency of the out reaction, 8×10^{-7} per HAT^r cell plated, is approximately 30% that of the in reaction. This frequency is 40-fold higher than the spontaneous mutation rate at the normal *HPRT* locus. Of the 6-TG^r colonies isolated, 88% had accurately excised the target vector sequences from the genome.

The mechanism of the out reaction, the excision event which generates 6-TG^r colonies, is likely to be either intrachromatid recombination between the 5-kb duplicated regions resulting from the insertion event (12–14) or an unequal sister chromatid recombination event between these same regions (19, 22). Both modes are depicted in Fig. 2. Our current information does not allow us to distinguish between them. However, it is of considerable interest that 95% of the 6-TG^r colonies examined from the HAT^r lines carrying the 4-bp insertion retained this insertion in the final locus. Because the modification is located 2 kb from the 5′ end of the duplicated region and 3 kb from the 3′ end, one would expect a 3:2 ratio in favor of losing the insertion if a simple crossover, occurring randomly along the length of the repeat, were responsible for the excision event. As this ratio is not observed, the 2-kb region may contain a hot spot for recombination.

Thompson et al. (25) have also examined reversion of an $HPRT^+$ targeted recombinant cell line that was generated in much the same way as ours. They similarly found that the target vector sequences had been excised from the genome in those cells which spontaneously reverted to the $hprt^-$ phenotype. Intrachromosomal recombination events occurring between direct repeats generated by gene targeting at the immunoglobulin Mu heavy-chain locus have been examined by Baker (1). The downstream repeat contained a 2-bp deletion relative to the upstream region. The majority of the recombinants were found to be gene conversion products; the excision products that did arise contained the longer (upstream) and shorter (downstream) repeats with equal frequency. This observation probably indicates that the bias we observed toward retaining the 4-bp insertion is gene specific rather than genome wide.

It has been shown that metabolic cross-feeding can interfere with 6-TG selections (9); therefore, the reversion frequency that we determined may be an underestimate. Because the ES cells always form aggregates upon plating, we could not eliminate the possibility of such cross-feeding. When using the in-out targeting procedure in other cell lines, it may be useful to plate HAT^r cells at a lower density to minimize the potential of revertant loss due to such cross-feeding.

An important part of our protocol (as described in Materials and Methods) was that the ES cells were not grown on feeder layers. Instead, the pluripotential nature of the cells was retained by supplementing the growth medium with purified human leukemia inhibitory factor (18, 26). This modification greatly simplifies the two-step selection procedure. Otherwise, the ES cells would have to be grown successively on two different feeder layers possessing opposite *HPRT* genotypes, requiring an adaptation stage between the selections. Also, 6-TG selection occurs slowly, requiring feeder-grown cultures to be passaged. We were able to simplify the selection procedure by eliminating this passage requirement. We were also able to isolate single clonal populations of the revertants, allowing for a more accurate determination of the reversion frequency.

The rationale for performing our experiments was to determine the feasibility of using an in-out targeting procedure to modify the mammalian genome. We have shown that both the integration and excision events occur accurately and at frequencies sufficient for use in a two-step targeting technique. Although we tested the two-step procedure in the directly selectable *HPRT* locus, the same procedure should be adaptable to modify nonselectable loci in an $hprt^-$ cell line by using the *HPRT* minigene described by Reid et al. (16).

The minigene would be carried on an integrating targeting vector, thereby allowing selection to be used for both the integration and excision events. Homologous recombinants are likely to be found after the in step at a frequency of 1 in 1,000 HAT^r cells, this being the ratio of transformed to targeted cells reported previously (20, 24). The targeted cell lines can then be identified by the polymerase chain reaction (see reference 27, for example). Including a small gap in the region of homology on the insertional vector provides a convenient primer binding site, since all gaps are repaired during the homologous insertion event (25a). Excision-derived $hprt^-$ revertants are likely to be found after the out step at a frequency of nearly 1 in 10^6 per targeted cell line.

The two-step selection scheme permits an enrichment for homologous recombinants carrying the desired modification, but the selectable sequences are not present in the final product. This in-out targeting procedure has the potential to be used in a variety of situations to create subtle modifications of the mammalian genome, a necessary prerequisite for fine-structure genetic analyses in mammalian systems.

ACKNOWLEDGMENTS

We thank Robert Schiestl for helpful discussions, Nick Gough for supplying leukemia inhibitory factor, Denise Dolenak and Hyung-

Suk Kim for performing preliminary experiments, and Sylvia Hiller for help with the Southern blots. We also thank Robert Schiestl, Tom Petes, and members of our laboratory group for critical readings of the manuscript.

This work was supported by grants GM20069 and HL37001 from the National Institutes of Health.

REFERENCES

1. **Baker, M. D.** 1989. High-frequency homologous recombination between duplicate chromosomal immunoglobulin Mu heavy-chain constant regions. Mol. Cell. Biol. 9:5500–5507.
2. **Boggs, S. S., R. G. Gregg, N. Borenstein, and O. Smithies.** 1986. Efficient transformation and frequent single-site, single-copy insertion of DNA can be obtained in mouse erythroleukemia cells transformed by electroporation. Exp. Hematol. 14:988–994.
3. **Caskey, C. T., and G. D. Kruh.** 1979. The HPRT locus. Cell 16:1–9.
4. **DeChiara, T. M., A. Efstratiadis, and E. J. Robertson.** 1990. A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting. Nature (London) 345:78–80.
5. **Doetschman, T. C., H. Eistetter, M. Katz, W. Schmidt, and R. Kemler.** 1985. The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. J. Embryol. Exp. Morphol. 87:27–45.
6. **Doetschman, T., R. G. Gregg, N. Maeda, M. L. Hooper, D. W. Melton, S. Thompson, and O. Smithies.** 1987. Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. Nature (London) 330:576–578.
7. **Doetschman, T., N. Maeda, and O. Smithies.** 1988. Targeted mutation of the HPRT gene in mouse embryonic stem cells. Proc. Natl. Acad. Sci. USA 85:8583–8587.
8. **Hooper, M., K. Hardy, A. Handyside, S. Hunter, and M. Monk.** 1987. HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. Nature (London) 326:292–295.
9. **Hooper, M. L., and J. H. Subak-Sharpe.** 1981. Metabolic cooperation between cells. Int. Rev. Cytol. 69:45–104.
10. **Jasin, M., and P. Berg.** 1988. Homologous integration in mammalian cells without target gene selection. Genes Dev. 2:1353–1363.
11. **Joyner, A. L., W. C. Skarnes, and J. Rossant.** 1989. Production of a mutation in mouse En-2 gene by homologous recombination in embryonic stem cells. Nature (London) 338:153–156.
12. **Liskay, R. M., A. Letsou, and J. L. Stachelek.** 1987. Homology requirements for efficient gene conversion between duplicated chromosomal sequences in mammalian cells. Genetics 115:161–167.
13. **Liskay, R. M., and J. L. Stachelek.** 1983. Evidence for intrachromosomal gene conversion in cultured mouse cells. Cell 35:157–165.
14. **Liskay, R. M., J. L. Stachelek, and A. Letsou.** 1984. Homologous recombination between repeated chromosomal sequences in mouse cells. Cold Spring Harbor Symp. Quant. Biol. 49:183–189.
15. **Mansour, S. L., K. R. Thomas, and M. R. Capecchi.** 1988. Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. Nature (London) 336:348–352.
16. **Reid, L. H., R. G. Gregg, O. Smithies, and B. H. Kohler.** 1990. Regulatory elements in the introns of the HPRT gene are necessary for its expression in mouse embryonic stem cells. Proc. Natl. Acad. Sci. USA 87:4299–4303.
17. **Scherer, S., and R. W. Davis.** 1979. Replacement of chromosome segments with altered DNA sequences constructed in vitro. Proc. Natl. Acad. Sci. USA 76:4951–4955.
18. **Smith, A. G., J. K. Heath, D. D. Donaldson, G. G. Wong, J. Moreau, M. Stahl, and D. Rogers.** 1988. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature (London) 336:688–690.
19. **Smithies, O., G. E. Connell, and G. H. Dixon.** 1962. Chromosomal rearrangements and the evolution of haptoglobin genes. Nature (London) 196:232–236.
20. **Smithies, O., R. G. Gregg, S. S. Boggs, M. A. Koralewski, and R. S. Kucherlapati.** 1985. Insertion of DNA sequences into the human chromosomal B-globin locus by homologous recombination. Nature (London) 317:230–234.
21. **Steeg, C. M., J. Ellis, and A. Bernstein.** 1990. Introduction of specific point mutations into RNA polymerase II by gene targeting in mouse embryonic stem cells: evidence for a DNA mismatch repair mechanism. Proc. Natl. Acad. Sci. USA 87:4680–4684.
22. **Sturtevant, A. H.** 1925. The effects of unequal crossing over at the Bar locus in Drosophila. Genetics 10:117–147.
23. **Thomas, K. R., and M. R. Capecchi.** 1987. Site-directed mutagenesis by gene-targeting in mouse embryo-derived stem cells. Cell 51:503–512.
24. **Thomas, K. R., K. R. Folger, and M. R. Capecchi.** 1986. High frequency targeting of genes to specific sites in the mammalian genome. Cell 44:419–428.
25. **Thompson, S., A. R. Clark, A. M. Pow, M. L. Hooper, and D. W. Melton.** 1989. Germline transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. Cell 56:313–321.

25a. **Valancius, V., and O. Smithies.** Unpublished data.

26. **Williams, R. L., D. J. Hilton, S. Pease, T. A. Willson, C. L. Stewart, D. P. Gearing, E. F. Wagner, D. Metcalf, N. A. Nicola, and N. M. Gough.** 1988. Myeloid leukemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature (London) 336:684–687.
27. **Zimmer, A., and P. Gruss.** 1989. Production of chimaeric mice containing embryonic stem (ES) cells carrying a homoeobox Hox 1.1 allele mutated by homologous recombination. Nature (London) 338:150–153.

Exhibit 4

MOLECULAR AND CELLULAR BIOLOGY, Nov. 1991, p. 5586–5591 Vol. 11, No. 11
0270-7306/91/115586-06$02.00/0

The Length of Homology Required for Gene Targeting in Embryonic Stem Cells

PAUL HASTY, JAIME RIVERA-PÉREZ, AND ALLAN BRADLEY*

Institute for Molecular Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030

Received 24 May 1991/Accepted 20 August 1991

Homologous recombination has been used to introduce site-specific mutations into murine embryonic stem (ES) cells with both insertion and replacement vectors. In this study, we compared the frequency of gene targeting with various lengths of homology and found a dramatic increase in targeting with an increase in homology from 1.3 to 6.8 kb. We examined in detail the relationship between the length of homology and the gene-targeting frequency for replacement vectors and found that a critical length of homology is needed for targeting. Adding greater lengths of homology to this critical length has less of an effect on the targeting frequency. We also analyzed the lengths of homology necessary on both arms of the vector for gene replacement events and found that 472 bp of homology is used as efficiently as 1.2 kb in the formation and resolution of crossover junctions.

Gene targeting, homologous recombination between introduced DNA and chromosomal DNA, has been used to make site-specific mutations in embryonic stem (ES) cells (3, 4, 6, 8, 9, 12, 14, 15). These mutations may subsequently be studied at the phenotypic level in chimeras and their progeny (9, 14). Despite the increasingly widespread use of gene-targeting schemes in ES cells, very little is known about the conditions that govern the frequency of gene targeting.

The essential feature of any gene-targeting vector is the sequence which is homologous to the target gene. Various lengths of homology may be important in the initial search, strand pairing, branch migration, and ultimate resolution of the recombination complex. The absolute length and distribution of homology in a targeting vector may have an impact on both the frequency of gene targeting and the integration pathway into the homologous target.

A previous investigation of extrachromosomal intramolecular recombination in mammalian cells illustrated a linear increase in the recombination frequency with the increased length of homology from 0.25 to 5 kb and a steep reduction in the frequency below 0.25 kb (11). These studies demonstrate that the recombination machinery may still function with limited homology, but they do not address vector/chromosome recombination which may be subject to different rate-limiting constraints. More recently, studies by Thomas and Capecchi have suggested an exponential relationship between target homology and targeting frequency with insertion and replacement vectors (15). However, another study observed a linear relationship between the length of homology and the targeting frequency (13).

We have described very different targeting frequencies between an insertion and replacement vector containing the same 6.8 kb of homologous sequences and alternative targeted integration patterns by replacement vectors (4). This suggests that aspects of vector configuration other than homology can greatly influence the frequency of the recombination event, and therefore the homology search may not be limiting. Targeting experiments at the *DHFR* locus with normal and amplified cell lines have also suggested that the homology search may not be limiting with 4.6 kb of targeting homology (17).

In the context of the hypoxanthine phosphoribosyltransferase (*hprt*) locus (10) in XY ES cells, all of the targeted clones can be isolated by selection for *hprt* inactivity in 6-thioguanine (TG). We investigated the length of vector homology that is required for efficient gene targeting with both insertion and replacement vectors and found a dramatic increase in the absolute targeting frequency when the length of homology was increased from 1.3 to 6.8 kb. A more detailed analysis of replacement vectors showed that homology of less than 1.7 kb was insufficient to generate targeted events. An additional 225 bp increased the frequency by at least fivefold. Homology increases from 1.9 to 4.2 kb and from 4.2 to 6.0 kb resulted in 16- and 3-fold, respectively, increases in the targeting frequency. We have also determined that the targeting of replacement vectors with just 472-bp sequence homology on the short arm was as efficient as 1.2 kb for a double-crossover (gene replacement) event.

MATERIALS AND METHODS

Construction of vectors. (i) Insertion vectors. IV1.3 is an *hprt* *Eco*RI fragment which contains exon 3. IV6.8 is a *Sac*I-*Eco*RI fragment which overlaps the IV1.3 homology and contains exons 2 and 3. MC1neopA (Stratagene) was inserted into the pTZ (Pharmacia) polylinker.

(ii) Replacement vectors that contain only exon 3. RV1.3 is an *Eco*RI fragment, RV1.7 is a *Hin*dIII-*Eco*RI fragment, and RV1.9 is an *Acc*I-*Eco*RI fragment.

(iii) Replacement vectors that contain exons 2 and 3. RV4.2 is a *Pvu*II-*Eco*RI fragment, RV6.0 is a *Bam*HI-*Eco*RI fragment, RV6.1 is a *Sac*I-*Acc*I fragment, and RV6.8 is a *Sac*I-*Eco*RI fragment. MC1neopA was inserted into the *Xho*I site of exon 3. RV1.3PGK and RV6.8PGK are the same as RV1.3 and RV6.8 except that PGKneobpA (14) was inserted into the *Xho*I site in exon 3.

Electroporations and tissue culture. The procedures have been described elsewhere (4, 9). Briefly, the DNA for electroporations was prepared by Triton lysis and banded once in CsCl. It was then linearized with the specified restriction enzyme and purified. The insertion vectors were

* Corresponding author.

linearized in the homology (*Xho*I). RV1.3 was linearized in the polylinker on the 5′ side of homology (*Sac*I). RV1.7 was linearized in the polylinker on the 5′ side of homology (*Sal*I) or in the polylinker on the 3′ side of homology (*Sac*I). RV1.9, RV4.2, and RV6.0 were linearized in the polylinker on the 3′ side of homology (*Sal*I). RV6.1 and RV6.8 were linearized at the 5′ edge of homology (*Sac*I) or in the polylinker on the 3′ side (*Sal*I). RV1.3PGK was linearized in the polylinker on the 5′ side of homology (*Sal*I), and RV6.8PGK was linearized in the polylinker on the 3′ side of homology (*Sal*I). The location of the linearization site, on either the 5′ or 3′ side of homology, did not alter the targeting frequency (4). The linearized vector was compared with uncut vector by gel electrophoresis migration to ensure the vector was completely cut. Electroporations were performed on AB1 (9) cells resuspended in phosphate-buffered saline at a density of 10^7 cells per ml. DNA (25 μg) was electroporated in 1 ml at 575 V/cm and 500 μF with a Bio-Rad Gene Pulser. The same microgram quantity was electroporated for each vector; therefore, the molar quantity differs by as much as a factor of 2. Only the absolute frequencies are affected by this molar difference. The absolute frequencies in Tables 1 and 2 are not adjusted to compensate for the molar difference; however, comparisons of absolute transfection and targeting frequencies between vectors of different size are adjusted in the text (there is a linear increase in transfection with an increase in the DNA concentration with in this range [2]). After each electroporation, the DNA and cells were incubated for 10 to 20 min at room temperature and then 10^7 cells were plated onto a 9-cm SNL76/7 feeder plate (9).

Three different methods were used to determine G418^r and TG^r as described elsewhere (4). In all three methods, the cells were not replated before selection such that each clone represents a single event. Briefly, (i) 24 h after electroporation, 180 μg of G418 (active ingredient) per ml was added to each 9-cm plate and maintained for 10 days in order to select and count G418^r colonies. TG was then added to a final concentration of 10^{-5} M. The cells were maintained under selection for another 21 days, after which TG^r colonies were counted. (ii) Plates were initially selected in 180 μg of G418 (active ingredient) per ml, and after 5 or 6 days a final concentration of 10^{-5} M TG was added; selection was maintained for another 10 days. (iii) Cells were scored by the incorporation of [^{3}H]hypoxanthine. Transfected cells were selected in G418 for 10 days and then labeled for 24 h in 4 ml of medium containing 0.04 μCi of [^{3}H]hypoxanthine. Cells that contained a functional *hprt* incorporated the labeled nucleoside and stained intensely black after fixation and exposure to a photographic emulsion. White colonies were scored as *hprt* negative.

All three methods were used for RV1.3 and RV6.8, and all three methods gave the same frequency of *hprt* inactivity for both constructs. For the other constructs, method 2 was used for 80 to 90% of the plates and method 1 was used for 10 to 20% of the plates. Methods 1 and 2 always gave the same absolute targeting frequency for each construct.

Southern blot analysis of targeted colonies. A *Bam*HI digest of genomic DNA was hybridized to a probe internal to the vector, *hprt* exons 2 and 3 (cDNA cut with *Hin*cII and *Hpa*II [10]); an *Nco*I digest of genomic DNA was hybridized to a 3′ external probe not present in the vector (a 0.5-kb fragment cut with *Pst*I and *Hin*dIII just 5′ to exon 4). *Bam*HI-digested DNA (5 μg) was separated by electrophoresis through a 0.7% agarose gel for 18 h at 60 V in 0.04 M Tris-acetate–0.001 M EDTA (TAE) buffer. The *Nco*I-digested DNA was

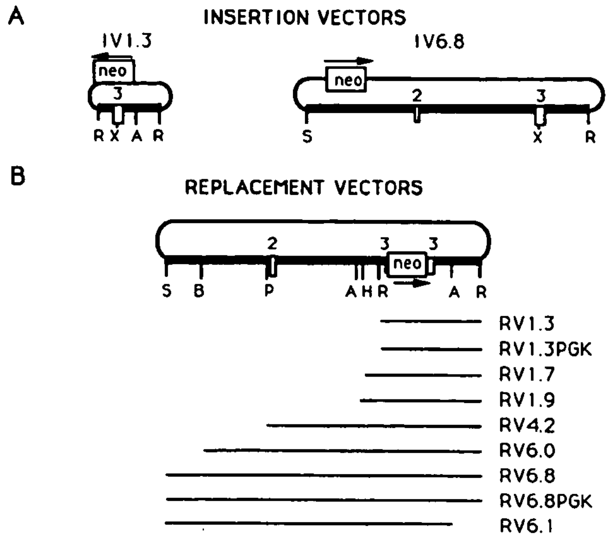


FIG. 1. *hprt* targeting vectors. The thick line is *hprt* homology in the vector, the thin line is plasmid, the box is the *neo* cassette (MC1neopA for all vectors except RV1.3PGK and RV6.8PGK, in which it is PGKneobpA), and the arrow is the direction of transcription. Exons 2 and 3 are labeled rectangles. Length of DNA is in kilobases. Restriction sites are *Acc*I (A), *Bam*HI (B), *Eco*RI (R), *Hin*dIII (H), *Pvu*II (P), *Sac*I (S), and *Xho*I (X). (A) Insertion vectors IV1.3 and IV6.8; (B) replacement vectors RV1.3, RV1.3PGK, RV1.7, RV1.9, RV4.2, RV6.0, RV6.8, RV6.8PGK, and RV6.1.

separated by electrophoresis through a 0.7% gel in TAE buffer for 24 h, using field inversion gel electrophoresis at 4°C (the buffer was circulated). DNA was transferred onto a GeneScreen Plus filter (Dupont) and hybridized to the specified probe under standard conditions. The probe was labeled by the random primed method, using the conditions specified by the manufacturer (Boehringer Mannheim).

RESULTS

Insertion vectors: a 250-fold difference in the targeting frequency with vectors of 1.3 and 6.8 kb. The insertion vectors IV1.3 and IV6.8 (Fig. 1A) were tested for their gene-targeting frequency. IV1.3 contains 1.3 kb of *hprt* DNA, and IV6.8 also contains the 5′ contiguous 5.5-kb *Eco*RI fragment of *hprt* DNA. Both constructs contain a *neo* expression cassette (MC1neopA), outside the *hprt* homology. The constructs were cut within the *hprt* homology at the *Xho*I site in exon 3 and electroporated into ES cells. The transfected ES cells were selected for G418 and TG resistance. The relative targeting frequencies are presented in Table 1.

Southern analysis of the TG^r colonies generated with IV1.3 confirmed that the vector had inserted into the *hprt* locus (Fig. 2). For these clones, the targeted events generated a 1.3-kb duplication of *hprt* homology separated by the *neo* gene and plasmid sequences. A *Bam*HI digest hybridized to exons 2 and 3 (exon 2 is external to the vector and exon 3 is internal to the vector) revealed that the 7.0-kb wild-type fragment was altered to fragments of 10.1 and 2.3 kb, as expected. An *Nco*I digest hybridized to a 3′ external probe reduced the 23-kb wild-type fragment to 14 kb as a result of the introduction of the *Nco*I site in MC1neopA. Previous analysis on TG^r clones generated with IV6.8 has

TABLE 1. Insertion vectors which target the *hprt* locus

Vector	Homology (kb)		Linear site	Total electroporations[a]	Total G418[r]	G418[r]/electroporation	Total *hprt*[−]	*hprt*[−]/G418[r]
	5'	3'						
IV1.3	0.132	1.2	*Xho*I	24	1,609	67	3	1/536
IV1.3	0.604	0.7	*Acc*I	10	2,055	205	1	1/2,055
IV6.8	5.6	1.2	*Xho*I	66	47,353	717	1,006	1/47

[a] Conditions for electroporation: 10^7 cells, 25 μg of DNA per ml.

demonstrated similar predicted targeted alleles generated by vector insertion in 19 of 20 TG[r] colonies analyzed (4).

The increase in homology from 1.3 to 6.8 kb increased the absolute targeting frequency by at least 250-fold and the relative targeting frequency by 11-fold (Table 1). The relative targeting frequency for IV1.3 is artificially high compared with that of IV6.8 due to poor expression of MC1neopA in IV1.3. MC1neopA expression is known to be sensitive to position effects (9, 14). Sequences in intron 2 of the murine *hprt* gene enhance expression of MC1neopA, which increases the background of G418[r] colonies for IV6.8 by 21-fold. The low targeting frequency for IV1.3 may be due to rate-limiting steps in the homology search such that 1.3 kb of homology may be inadequate to find the target locus or

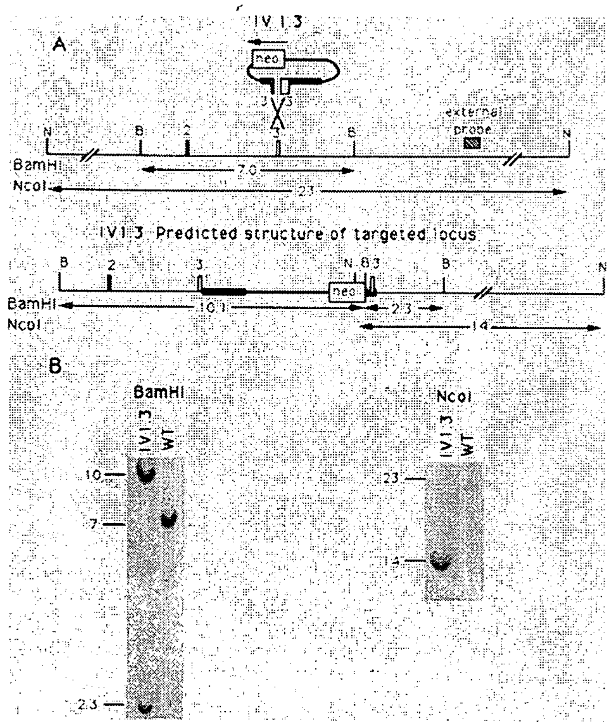


FIG. 2. Southern blot analysis for IV1.3. The thick line is *hprt* DNA of vector origin, the thin line is *hprt* DNA of genomic origin, and the line of intermediate thickness is plasmid. MC1neopA is a box labeled *neo*. Exons 2 and 3 are numbered rectangles and serve as a probe (taken from cDNA) for the *Bam*HI digest. The striped box is the external probe for the *Nco*I digest. The length of DNA is in kilobases. Shown are fragments and predicted size fragments for single-reciprocal recombination seen with a *Bam*HI (B) digest hybridized to exons 2 and 3 and an *Nco*I (N) digest hybridized to the external probe. WT, wild type.

assemble a recombination complex. Alternatively, there was just 132 bp of homology on the 5' end, which may be too little homology on one end for efficient homologous recombination.

IV1.3 targets inefficiently because of a lack of total homology. The inefficient targeting of IV1.3 may be due to a lack of total homology (1.3 kb) or too little homology on the 5' end (132 bp). To test whether the distribution of homology or total homology was responsible for the poor targeting efficiency, IV1.3 was linearized near the center of homology with *Acc*I, leaving homologous arms of 604 and 700 bp (Fig. 1A). IV1.3 cut with *Acc*I generated one TG[r] colony in 10^8 cells electroporated (Table 1). IV1.3 cut with either *Xho*I or *Acc*I generated targeted colonies at a similar absolute frequency. This indicated that 1.3 kb of total homology was inadequate to target *hprt* efficiently. Presumably, the limited homology of 1.3 kb becomes rate limiting in the targeting pathway because of the search for the target sequences or formation of recombination intermediates.

Replacement vectors: a 190-fold difference in targeting efficiency with vectors of 1.3 and 6.8 kb. Replacement vectors (Fig. 1B) RV1.3 and RV6.8 (4) were tested for their gene-targeting frequency at the *hprt* locus. RV1.3 and RV6.8 contain the same *hprt* fragments as do IV1.3 and IV6.8, respectively. The MC1neopA expression cassette was inserted inside *hprt* exon 3 at the *Xho*I site. Prior to transfection, these vectors were linearized by cutting outside the targeting homology. ES cells were selected in G418 and TG. RV1.3 did not yield any TG[r] colonies in 34 × 10^7 cells electroporated (Table 2). RV6.8 generated TG[r] colonies at an absolute frequency of 2.9 TG[r] colonies per 10^7 cells electroporated. We have previously shown that only a fraction of these TG[r] targeted clones have the predicted gene replacement event (4).

An increase in homology from 1.3 to 6.8 kb increased the targeting frequency. There was at least a 190-fold difference in the absolute targeting frequency between RV1.3 and RV6.8. We recovered TG[r] targeted clones with IV1.3 but not with RV1.3. The reason for this may reflect either our previous observations of the significantly higher targeting frequency with insertion vectors (4) or the possibility that low transfection efficiency of RV1.3 prohibits targeting.

To determine the effect of the transfection frequency on the targeting frequency, the MC1neopA selection cassette was replaced with the position-independent cassette PGK neobpA (14) in both RV1.3 and RV6.8 to make RV1.3PGK and RV6.8PGK (4) (Fig. 1B). Unlike RV1.3 and RV6.8, RV1.3PGK and RV6.8PGK have similar transfection efficiencies (Table 2). Despite the high transfection frequency for RV1.3PGK, there were no TG[r] colonies generated out of 46,144 G418[r] colonies. The absolute targeting frequencies for RV6.8PGK were the same as for RV6.8 (Table 2). Southern analysis for clones targeted with RV6.8PGK have been described previously (4).

TABLE 2. Replacement vectors which target the *hprt* locus

Vector	Homology (kb)		Total electroporations	Total G418[r]	G418[r]/electroporation[a]	Total *hprt*[−]	*hprt*[−]/electroporation	*hprt*[−]/G418[r]	Integration pattern[b]				
	5′	3′							Vector insertion: Long arm	Vector insertion: Short arm	Gene replacement	VIM	UD
RV1.3	0.132	1.2	34	3,752	110	0							
RV1.3PGK	0.132	1.2	15	46,144	3,076	0							
RV1.7	0.536	1.2	28	4,057	145	0							
RV1.9	0.761	1.2	51	32,041	628	7	0.14	1/4,577	0	1	1	0	0
RV4.2	3.0	1.2	28	43,390	1,550	43	1.5	1/1,009	3	0	0	2	0
RV6.0	4.8	1.2	28	51,758	1,848	107	3.8	1/483	1	0	2	2	0
RV6.8	5.6	1.2	22	17,846	811	64	2.9	1/279	12	0	1	3	1
RV6.8PGK	5.6	1.2	21	28,702	1,367	58	2.8	1/495	5	0	2	1	1
RV6.1	5.6	0.472	29	15,547	536	37	1.5	1/420	12	1	7	3	1

[a] Conditions for electroporation: 10^7 cells, 25 μg of DNA per ml.

[b] Vector insertion, integration of the entire vector, including the bacterial plasmid, with crossover junctions on only the long or short arm (4); gene replacement, a nonreciprocal exchange of information that replaces the target locus with only the mutation, the bacterial plasmid is not included (4, 15); VIM, vector integration modified, the integration pattern of the vector is not understood (4); UD, undetermined, the integration of the targeting vector could not be verified.

Replacement vectors: the influence of total homology on the targeting frequency. An overlapping set of replacement vectors was made to determine the relationship between the total length of homology and the frequency of homologous recombination. We have progressively removed 5′ *hprt* homology from RV6.8 to make replacement vectors with 1.7 (RV1.7), 1.9 (RV1.9), 4.2 (RV4.2), and 6.0 (RV6.0) kb of *hprt* homology (Fig. 1B). The vectors were linearized outside the homology prior to transfection. The targeting frequencies are presented in Table 2 and Fig. 3. RV1.7 did not generate any TG[r] colonies in 36 × 10^7 cells electroporated; RV1.9, RV4.2, and RV6.0 generated TG[r] colonies at absolute frequencies of 0.14 × 10^{-7}, 1.5 × 10^{-7}, and 3.8 × 10^{-7} cells, respectively.

Southern analysis of TG[r] colonies generated with RV1.9, RV4.2, and RV6.0 showed the same alternative integration pathways that we described previously (4) (data not shown). In total, there were three gene replacement events (a nonreciprocal exchange that occurs via double-reciprocal recombination or gene conversion), five vector integration events

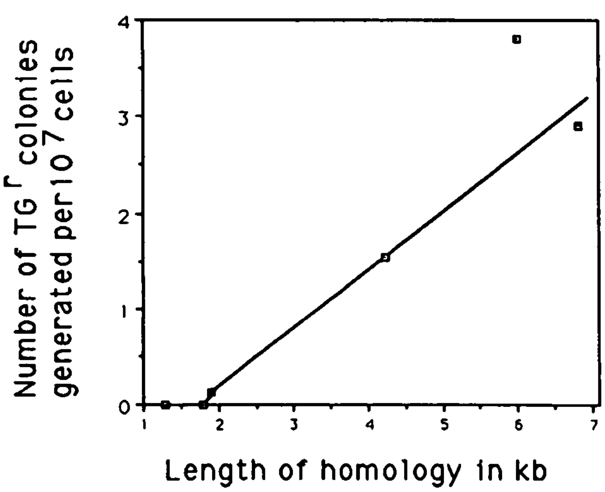


FIG. 3. Absolute frequency of recombination plotted against the length of homology to the target sequence with replacement vectors.

(integration of the entire vector, including the bacterial plasmid), and four modified integration events in which the vector integrated into the target locus in an undefined manner.

An increase of *hprt* homology from 1.3 to 1.7 kb did not result in the detectable generation of targeted colonies; however, a further increase of just 225 bp to 1.9 kb did generate targeted colonies. An addition of *hprt* homology from 1.9 to 4.2 kb resulted in a 16-fold elevation in the targeting frequency, and a further increase of homology from 4.2 to 6.0 kb resulted in an additional 3-fold increase in targeting. Greater homology did not increase the targeting frequency (Fig. 3).

Crossover junctions can utilize just 472 bp of homology in gene replacement events. One explanation for the low targeting frequency of vectors such as RV1.3 and RV1.7 is that the amount of homology on one vector arm is too short to form a productive recombination complex. This step would then become rate limiting in the formation of the final recombination product. To test this, we constructed RV6.1, which has just 0.472 kb on the short arm, 64 bp less than RV1.7, which did not generate targeted colonies at a detectable frequency. RV6.1, linearized outside the homology, generated TG[r] colonies at a frequency of 1.3 × 10^{-7} cells electroporated. This is a modest decrease in the targeting frequency compared with that of RV6.8, which has 1.2 kb on the short arm. It was important to verify that the 472-bp short arm was utilized for crossovers because the majority of targeted clones do not form crossover junctions on the short arm of homology for RV6.8 (4).

Southern analysis on 24 TG[r] colonies generated with RV6.1 (Fig. 4) determined that 7 were targeted by a gene replacement event (Fig. 4B, clones 9, 10, 12, and 13) and 12 were targeted by a vector insertion event on the long arm (Fig. 4B, clones 7, 11, and 15), 1 was targeted by a vector insertion event on the short arm (Fig. 4B, clone 8), and 3 were targeted by a modified integration event (Fig. 4B, clone 14). A *Bam*HI digest hybridized to an internal probe (exons 2 plus 3) revealed that the 7.0-kb wild-type band was altered to 5.9 and 2.3 kb for gene replacement events, 7.0, 5.9, and 4.2 kb for vector integration events that utilize the long arm, and 9.8, 6.6, and 2.3 kb for the vector integration events that utilize the short arm. An *Nco*I digest hybridized to an

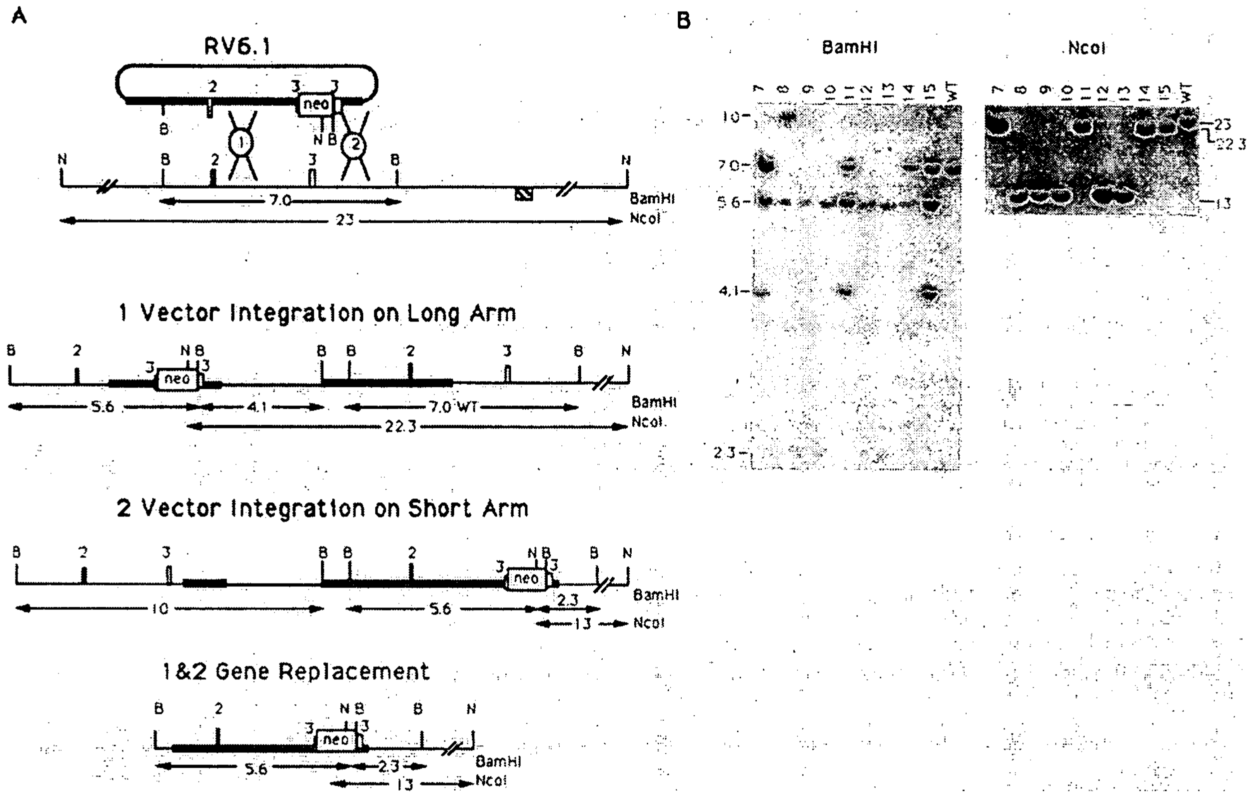


FIG. 4. Southern blot analysis for RV6.1. The thick line is *hprt* DNA of vector origin, the thin line is *hprt* DNA of genomic origin, and the line of intermediate thickness is plasmid. MC1neopA is a box labeled *neo*. Exons 2 and 3 are numbered rectangles and serve as the internal probe (taken from cDNA). The striped box is the external probe. The encircled number 1 represents the crossover point on the long arm, and the encircled number 2 represents the crossover point on the short arm. The length of DNA is in kilobases. Shown are predicted size fragments seen with *Bam*HI (B) digests hybridized to exons 2 and 3 and *Nco*I (N) digests hybridized to the external probe. WT, wild type.

external probe altered the 23-kb wild-type band to 13 kb for gene replacement events and the vector integration events on the short arm and 22.3 kb for vector integration events on the long arm. The modified integration event seen in clone 14 in Fig. 4B appears to be a vector integration event on the long arm on the *Nco*I digest. The *Bam*HI digest shows the 4.1-kb band to be absent and a 3.7-kb band to be present. This may have occurred as a result of exonuclease activity associated with the end-to-end joining prior to the integration (4, 16). The other two modified integration events also appear to be vector integrations on the long arm in association with exonuclease activity (data not shown).

These results demonstrate that just 472 bp of sequence homology is adequate to generate a functional crossover event. This implies that the rate-limiting step that reduces the gene targeting frequency for RV1.3 and RV1.7 to below detectable limits is the total length of homology, not the formation and resolution of the crossover junction on the short arm of homology.

DISCUSSION

The gene-targeting frequency is dependent on the length of the target homology. We have shown there is at least a 200-fold increase in gene-targeting efficiency associated with an increase of homology from 1.3 to 6.8 kb for replacement and insertion vectors.

A more detailed analysis of replacement vectors was unable to detect targeting in a total of 3.4×10^8 cells electroporated with 1.3 kb of homology and 2.8×10^8 cells electroporated with 1.7 kb of homology. This low targeting frequency was not associated with the low transfection efficiency of the vectors, since the 1.3 kb of homology, modified to contain a *neo* cassette subject to fewer positional effects and with a stronger promoter, did not yield TG^r colonies in 46×10^3 G418^r cells analyzed. Within the ranges of homology from 1.7 to 1.9, 1.9 to 4.2, and 4.2 to 6.0 kb, there are increases in the targeting frequency of >21-fold, 7-fold, and 1.6-fold per kb, respectively. The largest change in the targeting efficiency compared with a proportional change in homology was the >21-fold per kb seen with the 225-bp increase from RV1.7 to RV1.9. It is possible that the added 225 bp of homology contains a hot spot for recombination. Previously reported data showed that a 2-fold increase in homology resulted in a 20-fold increase in the targeting frequency (15). These added lengths of homology may also have incorporated hot spots for gene targeting. However, a more likely explanation is that a critical length of homology is required for targeting, and once this requirement is met, then further increases in homology have less effect. This has been observed in extrachromosomal recombination in mammalian cells (1, 11). RV1.3 and RV1.7 may not target efficiently because of inadequate total homology or inadequate homology on the short arm.

The total length of homology has a greater impact on the targeting efficiency than does the length of homology on the short arm. Colonies targeted with RV6.1, which only has 472 bp on the short arm, were analyzed to map the length of homology necessary to form and resolve a Holliday (5) junction or initiate and resolve a gene conversion event. The 472-bp short arm was shown to be utilized as efficiently as the 1.2-kb short arm in RV6.8 (both vectors have 5.6 kb of homology on the long arm). This implies that the rate-limiting step that reduces the gene-targeting frequency for RV1.7 (short arm of 536 bp) to below detectable limits is the inadequate total homology and not the amount on the short arm. We have also shown that the targeting frequency of IV1.3 does not detectably change when the vector is cut more centrally.

Replacement vectors usually integrate the entire vector into the target locus, including the bacterial plasmid, with the crossover occurring on the long arm (4). The bias in crossover formation with respect to the long arm indicates a relationship between homology and the formation of a crossover. However, 472 bp is sufficient and apparently not rate limiting for the formation and resolution of crossover junctions. This observation suggests that it is beneficial to have a greater length of homology to search for the target locus and form recombination intermediates than is absolutely necessary to productively recombine and form a crossover junction.

These data also support our previous observation on the difference in recombination frequencies with insertion and replacement vectors (4). Here the 1.3-kb insertion vector is significantly more efficient at targeting than is the 1.3-kb replacement vector. This observation indicates that in the context of the same amount of homologous sequences, the targeting mechanism can also be a rate-limiting step. The formation and resolution of recombination intermediates is probably more efficient with the two adjacent homologous ends presented by an insertion vector as has been demonstrated in yeast cells (7).

There are many facets that affect the frequency of gene targeting. In the context of *hprt*, we have illustrated the relationship between the targeting frequency and the length of homology in a targeting vector. We have previously demonstrated that the targeting frequency can be manipulated by the vector configuration (4). Targeting efficiency is also dependent on the location of the target in the genome. The appreciation of the relationship between vector design and the targeting frequency along with an understanding of the parameters that affect the availability of any locus for gene targeting will ultimately enable the targeting of any genomic region at high efficiency.

ACKNOWLEDGMENTS

We acknowledge J. D. Wallace for technical assistance and thank our colleagues Ann Davis, Marty Matzuk, Ramiro Ramírez-Solis, and Hui Zheng for critical review of the manuscript.

This work was supported by grants from the NIH (A.B.), the Searle Scholars Program/Chicago Community trust (A.B.), the Granada Corporation (P.H.), and the Cystic Fibrosis Foundation (A.B., P.H., and J.R.).

REFERENCES

1. **Ayares, D., L. Chekuri, K. Song, and R. Kucherlapati.** 1986. Sequence homology requirements for intermolecular recombination in mammalian cells. Proc. Natl. Acad. Sci. USA **83:**5199–5203.
2. **Davis, A. C.** Personal communication.
3. **Hasty, P., R. Ramirez-Solis, R. Krumlauf, and A. Bradley.** 1991. Introduction of a subtle mutation into the *Hox-2.6* locus in embryonic stem cells. Nature (London) **350:**243–246.
4. **Hasty, P., J. Rivera-Perez, C. Chang, and A. Bradley.** 1991. Target frequency and integration pattern for insertion and replacement vectors in embryonic stem cells. Mol. Cell. Biol. **11:**4509–4517.
5. **Holliday, R.** 1964. A mechanism for gene conversion in fungi. Genet. Res. **5:**282–304.
6. **Johnson, R. S., M. Sheng, M. E. Greenberg, R. D. Kolodner, V. E. Papaioannou, and B. M. Spiegelman.** 1989. Targeting of nonexpressed genes on embryonic stem cells via homologous recombination. Science **245:**1234–1236.
7. **Orr-Weaver, T. L., J. W. Szostak, and R. J. Rothstein.** 1981. Yeast transformation: a model system for the study of recombination. Proc. Natl. Acad. Sci. USA **78:**6354–6358.
8. **Mansour, S. L., K. R. Thomas, and M. R. Capecchi.** 1988. Disruption of the proto-oncogene *int-2* in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. Nature (London) **336:**348–352.
9. **McMahon, A. P., and A. Bradley.** 1990. The *Wnt-1* (*int-1*) proto-oncogene is required for development of a large region of the mouse brain. Cell **62:**1073–1085.
10. **Melton, D. W., D. S. Konecki, J. Brennand, and C. T. Caskey.** 1984. Structure, expression, and mutation of the hypoxanthine phosphoribosyltransferase gene. Proc. Natl. Acad. Sci. USA **81:**2147–2151.
11. **Rubnitz, J., and S. Subramani.** 1984. The minimum amount of homology required for homologous recombination in mammalian cells. Mol. Cell. Biol. **4:**2253–2258.
12. **Schwartzberg, P. L., S. P. Goff, and E. J. Robertson.** 1989. Germ-line transmission of a c-*abl* mutation produced by targeted gene disruption in ES cells. Science **246:**799–803.
13. **Shulman, M. J., L. Nissen, and C. Collins.** 1990. Homologous recombination in hybridoma cells: dependence on time and fragment length. Mol. Cell. Biol. **10:**4466–4472.
14. **Soriano, P., C. Montgomery, R. Geske, and A. Bradley.** 1991. Targeted disruption of the c-src proto-oncogene leads to osteopetrosis in mice. Cell **64:**693–702.
15. **Thomas, K. R., and M. R. Capecchi.** 1987. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell **51:**503–512.
16. **Wilson, J. H., P. B. Berget, and J. M. Pipas.** 1982. Somatic cells efficiently join unrelated DNA segments end to end. Mol. Cell. Biol. **2:**1258–1269.
17. **Zheng, H., and J. Wilson.** 1990. Gene targeting in normal and amplified cell lines. Nature (London) **344:**170–173.

Exhibit 5

MOLECULAR AND CELLULAR BIOLOGY, Sept. 1991, p. 4509–4517
0270-7306/91/094509-09$02.00/0

Vol. 11, No. 9

Target Frequency and Integration Pattern for Insertion and Replacement Vectors in Embryonic Stem Cells

PAUL HASTY, JAIME RIVERA-PÉREZ, CHRISTINE CHANG, AND ALLAN BRADLEY*
Institute for Molecular Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030

Received 31 December 1990/Accepted 24 June 1991

Gene targeting has been used to direct mutations into specific chromosomal loci in murine embryonic stem (ES) cells. The altered locus can be studied in vivo with chimeras and, if the mutated cells contribute to the germ line, in their offspring. Although homologous recombination is the basis for the widely used gene targeting techniques, to date, the mechanism of homologous recombination between a vector and the chromosomal target in mammalian cells is essentially unknown. Here we look at the nature of gene targeting in ES cells by comparing an insertion vector with replacement vectors that target *hprt*. We found that the insertion vector targeted up to ninefold more frequently than a replacement vector with the same length of homologous sequence. We also observed that the majority of clones targeted with replacement vectors did not recombine as predicted. Analysis of the recombinant structures showed that the external heterologous sequences were often incorporated into the target locus. This observation can be explained by either single reciprocal recombination (vector insertion) of a recircularized vector or double reciprocal recombination/gene conversion (gene replacement) of a vector concatemer. Thus, single reciprocal recombination of an insertion vector occurs 92-fold more frequently than double reciprocal recombination of a replacement vector with crossover junctions on both the long and short arms.

It is now possible to study the function of specific genes in the mouse by the application of gene targeting in embryonic stem (ES) cells. ES cells which have undergone a specific gene-targeting reaction may subsequently be used to establish the allele in the germ line of mice (2, 8, 12, 17, 20, 23, 27). However, the application of gene targeting is limited by random recombination, which is 30- to 40,000-fold more frequent than homologous recombination; thus, most transfected cells are not targeted (2–4, 6, 7, 9, 11, 12, 17, 20, 22–24). Understanding the parameters that influence the absolute and relative frequencies of gene targeting will have a direct impact on many experiments that require the gene to be mutated in ES cells.

Two distinct vector designs have been investigated in gene targeting experiments, replacement and insertion vectors. Replacement (omega type) vectors are intended to replace homologous gene sequences with a selectable mutation (22) and have been most commonly used to make null alleles in ES cells (2, 8, 12, 17, 20, 23, 27). Typically, replacement vectors contain sequences homologous to the target, with a *neo* expression cassette inserted within the region of homology so that it disrupts an exon and creates a mutation. The vector is linearized outside the region of homology. Following homologous recombination, sequences in the vector replace the genomic sequence by double reciprocal recombination or gene conversion, thereby inserting the *neo* mutation into the target gene (22). The nonhomologous sequences at the free ends of the construct are excised (11). Insertion vectors (O type) are the alternative vector type, which are intended to insert the entire construct into the homologous sequences (22). These vectors are cut at a site within the homology. A double-strand break in the homology has been shown to increase both the targeting frequency in yeast cells (15) and extrachromosomal recombination in mammalian cells (10, 19, 26). Insertion vectors have been used in ES cells for gene correction (3) and gene disruption (22) at the selectable *hprt* locus.

The most widely used vectors are replacement vectors. Their general acceptance is based upon a preliminary report that recombination frequencies were similar to those with insertion vectors (22) and that positive-negative selection is applicable to this vector type (11). Recently, some reports have indicated that nonpredicted integration events were occurring when replacement vectors were used (18, 23). The types of integration events that were occurring suggested that screens such as polymerase chain reaction (PCR) and negative selection may eliminate certain classes of targeted recombinants from analysis. A description of the entire spectrum of targeted integration events, free from any bias imposed by selection and screening techniques, will facilitate an understanding of the gene-targeting pathways in ES cells. This, in turn, will enable vectors to be designed that can most efficiently utilize the host recombination machinery to target genes.

In order to study the frequency and integration events of homologous recombination in ES cells, we have chosen to target a gene involved in the purine salvage pathway, that for hypoxanthine phosphoribosyltransferase (*hprt*) (13). *hprt* is located on the X chromosome and is a single-copy gene in XY ES cells; therefore, all targeted events regardless of fidelity can be selected by the absence of *hprt* function with a toxic base analog, 6-thioguanine (TG). We show that an insertion vector targeted up to ninefold more frequently than a replacement vector with the same homologous sequences and the same *neo* expression cassette. We also report that most of the clones targeted with replacement vectors were not generated by the predicted gene replacement event; rather, the entire construct was usually integrated into the *hprt* locus. The targeting efficiency and integration pattern were unaffected even if the targeted homology was released from the plasmid. Our observations demonstrate that circularized products and/or concatemers form before replacement vectors integrate into the target locus.

* Corresponding author.

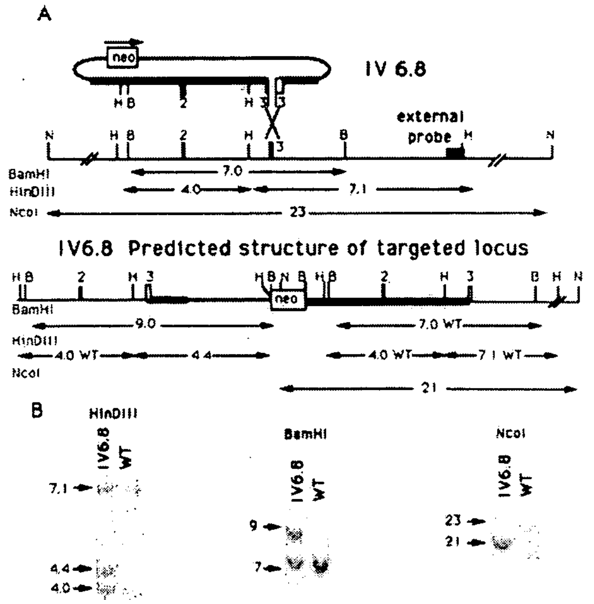


FIG. 1. Southern blot analysis of TG^r clones generated with IV6.8. (A) Wild-type (WT) fragments and predicted sizes of fragments for single reciprocal recombination seen with a *Bam*HI (B) digest, a *Hin*dIII (H) digest, and an *Nco*I (N) digest. The thick line is *hprt* DNA of vector origin, the thin line is *hprt* DNA of genomic origin, and the line of intermediate thickness is plasmid DNA. MClneopA is a box labeled *neo*. Exons 2 and 3 are numbered rectangles and serve as the internal probe (taken from cDNA). The filled box is the external probe. The length of DNA is in kilobases. (B) *Bam*HI and *Hin*dIII digests were hybridized to exons 2 and 3. The *Nco*I digest was hybridized to the external probe.

MATERIALS AND METHODS

Construction of vectors. The insertion vector IV6.8 is a *Sac*I-*Eco*RI fragment of the murine *hprt* gene that contains exons 2 and 3 inserted into pTZ (Pharmacia) at the *Sac*I and *Eco*RI sites and contains a *neo* expression cassette, MClneopA (Stratagene), in the polylinker. Replacement vectors RV6.8 and RV6.8PGK were made from the same homologous sequences as IV6.8; MClneopA or PGKneobpA (1, 16, 20), respectively, was inserted into the *Xho*I site of exon 3.

Electroporation and tissue culture. The DNA was prepared for electroporation by Triton lysis and banded once in CsCl. It was then linearized with the specified restriction enzyme, phenol-chloroform extracted, ethanol precipitated, and resuspended in 10 mM Tris (pH 7.5)–1 mM EDTA at a concentration between 0.5 and 1.0 μg/μl. To ensure complete digestion, the cut DNA was compared with uncut DNA by a mobility shift assay with gel electrophoresis. DNA was not electroporated unless it was completely cut. Electroporations were performed on AB1 (12) cells collected from confluent 9-cm plates of trypsin-treated cells about 3 h after refeeding. The AB1 cells were centrifuged for 5 min at 1,000 rpm and resuspended in phosphate-buffered saline (Ca^{2+} and Mg^{2+} free) at a density of 10^7/ml. DNA (25 μg) was electroporated in 1 ml at 575 V/cm and 500 μF with a Bio-Rad Gene Pulser. After each electroporation, the DNA and cells were incubated for 10 to 20 min at room temperature, and then 10^7 cells were plated onto a 9-cm SNL76/7 feeder plate (12) with 10 ml of medium (Dulbecco's modified Eagle's medium with 15% fetal calf serum, 2 mM glutamine, 30 μg of penicillin per ml, 50 μg of streptomycin per ml, and 10^{-4} M β-mercaptoethanol).

Three different methods were used to determine resistance to G418 and TG. Method 1: At 24 h after electroporation, 180 μg of G418 (active ingredient) per ml was added to each 9-cm plate and maintained for 10 days in order to select and count G418-resistant colonies. TG was then added to a final concentration of 10^{-5} M. The cells were maintained under selection for another 21 days, after which $G418^r$ and TG^r colonies were counted. Method 2: Plates were also selected in G418 and TG simultaneously at the concentrations described for method 1. The G418 was added 24 h after the electroporation, and then TG was added 5 to 6 days later. Double selection was maintained for another 10 days, at which point $G418^r$ and TG^r colonies were counted. Method 3: Cells were scored by the incorporation of [^{3}H]hypoxanthine. Transfected cells were selected in G418 for 10 days and then labeled for 24 h in 4 ml of medium containing 0.04 μCi of [^{3}H]hypoxanthine. The colonies were washed twice in phosphate-buffered saline and then fixed in methanol-acetic acid vapor (3:1) for 10 min. The colonies were extracted twice in ice-cold 10% trichloroacetic acid and washed in water overnight. The colonies were dried, and a thin layer of nuclear emulsion was applied. The emulsion was exposed in the dark for 3 days at 4°C. Silver grains were developed and fixed by standard procedures. Cells that contained a functional *hprt* gene incorporated the labeled nucleoside and were intensely black. White colonies were scored as *hprt* negative. All three methods were used for RV6.8 in order to establish the accuracy of each method. The same results were obtained for all three methods. For RV6.8PGK and IV6.8, method 1 was used on 10 to 20% of the plates to score $G418^r$ colonies and then $G418^r$ plus TG^r colonies, and method 2 was used on 80 to 90% of the plates to score $G418^r$ plus TG^r colonies.

Southern blot analysis of targeted colonies. A *Bam*HI digest and a *Hin*dIII digest of genomic DNA were hybridized to two probes internal to the vector, *hprt* exons 2 and 3 (cDNA cut with *Hin*cII and *Hpa*I [12]), and plasmid sequences; *Nco*I, *Nco*I-*Sal*I, and *Hin*dIII digests of genomic DNA were hybridized to a 3′ external probe not present in the vector (a 0.5-kb fragment cut with *Pst*I and *Hin*dIII just 5′ to exon 4). Isolated TG^r colonies were passaged once onto gelatinized plates without feeder cells to reduce feeder cell DNA. *Bam*HI-*Hin*dIII-digested DNA (5 μg) was separated by electrophoresis through a 0.7% agarose gel for 18 h at 60 V in TAE buffer. The *Nco*I and *Nco*I-*Sal*I-digested DNA was separated by electrophoresis through a 0.7% gel in TAE buffer for 24 h by field inversion gel electrophoresis at 4°C. The buffer was circulated. DNA was transferred onto a GeneScreen Plus filter (Du Pont) and hybridized to the specified probe under standard conditions. The probe was random primer labeled under the manufacturer's specified conditions (Boehringer Mannheim).

RESULTS

Insertion vectors target the *hprt* locus more efficiently than replacement vectors. To test the targeting efficiency of vectors with the same homologous sequences, we constructed insertion and replacement vectors which contain 6.8 kb of homologous sequence. The insertion vector IV6.8 contains the MClneopA expression cassette in the plasmid backbone (Fig. 1A), while the replacement vectors RV6.8 (see Fig. 3A) and RV6.8PGK (see Fig. 2A) contain MClneopA and PGK-neobpA, respectively, in the *Xho*I site of exon 3. IV6.8 was

TABLE 1. Insertion and replacement vectors which target the *hprt* locus

Vector	Restriction digest	No. of electro-porations	Total no. of clones		Ratio, hprt⁻/G418ʳ	No. of TGʳ clones analyzed	Integration pattern[a] (no. of TGʳ clones)					
			G418ʳ	*hprt*⁻			L1&Lx or VILA	S1&Sx or VISA	VIM	L1&S1	L1&Sx	UD
IV6.8	*Xho*I	46	23,884	610	1:39	20	ND	ND	1	ND	ND	—
RV6.8	*Sac*I	8	7,952	36	1:221	7	4	0	3	0	0	0
RV6.8	*Sal*I	10	9,894	28	1:353	10	8	0	0	0	1	1
RV6.8	*Sac*I + *Sal*I	29	15,530	59	1:263	14	8	0	3	0	0	3
RV6.8PGK	*Sal*I	21	28,702	58	1:495	9	5	0	1	2	0	1

[a] VILA, vector insertion, long arm; VISA, vector insertion, short arm; VIM, vector integration, modified; L1, long arm of the first unit in a multimer; S1, short arm of the first unit in a multimer; Lx, long arm of unit x in a multimer; Sx, short arm of unit x in a multimer; UD, undetermined; ND, not determined.

linearized at the *Xho*I site in exon 3. RV6.8 was cut at the 5′ edge of homology with *Sac*I. Both RV6.8 and RV6.8PGK were cut at the 3′ edge of homology with *Sal*I, which leaves 30 bp of heterologous sequence on the 3′ end. The DNA was electroporated into ES cells, and the transfection events were determined by selection in G418. A second selection with TG was used to isolate clones without a functional *hprt* gene. IV6.8 targeted at a relative frequency of 1 TGʳ to 39 G418ʳ colonies (Table 1). RV6.8 cut with *Sac*I, RV6.8 cut with *Sal*I, and RV6.8PGK cut with *Sal*I yielded TGʳ/G418ʳ colony ratios of 1:221, 1:353, and 1:495, respectively (Table 1). PGKneobpA yields more G418ʳ colonies per electroporation than MC1neopA (20), and this accounts for the slightly lower targeting frequency seen with RV6.8PGK. The absolute targeting frequencies were the same. IV6.8 proved to be about five- to ninefold more efficient at gene targeting than RV6.8 cut with *Sac*I and *Sal*I, respectively.

IV6.8 integrates the entire vector into the *hprt* locus during homologous recombination. Targeted integration of IV6.8 will form a duplication of the 6.8 kb of *hprt* homology separated by the plasmid and *neo* sequences (Fig. 1). To confirm that the TGʳ clones had the predicted structure, *Bam*HI and *Hin*dIII digests were hybridized to an internal probe, *hprt* exons 2 and 3 from cDNA. A *Bam*HI digest on wild-type DNA gives a 7.0-kb fragment. The targeted insertion events should result in a duplication; the 7.0-kb wild-type fragments were unchanged, and the predicted additional 9.0-kb fragments were formed. A *Hin*dIII digest was hybridized to the same internal probe; as expected, the 7.1- and 4.0-kb wild-type fragments were retained and additional 4.4-kb fragments were generated. The novel fragments on the *Bam*HI and *Hin*dIII digests verified the presence of a 6.8-kb *hprt* duplication which contains exons 2 and 3. To ensure that the 6.8-kb duplication was separated by a plasmid sequence and a *neo* gene, the *Bam*HI and *Hin*dIII blots were stripped and hybridized to plasmid sequences. The 9.0-kb fragment on the *Bam*HI blot and the 4.4-kb fragment on the *Hin*dIII blot hybridized to the plasmid sequences, as predicted (data not shown). This analysis showed that 19 of 20 TGʳ clones had the predicted restriction pattern. Concatemers appeared to have integrated into the target locus in 3 of these 19 clones. One clone gave a restriction pattern that was not expected and could not be interpreted.

The restriction patterns seen for the *Bam*HI and *Hin*dIII digests would also be seen if a concatemer of the vector integrated randomly in the genome adjacent to *Bam*HI and *Hin*dIII sites at the appropriate distance in association with a spontaneous mutation of *hprt* in a location other than the targeting homology. It is unlikely for this to occur; however, in order to definitively determine that targeting occurred at *hprt*, a probe outside the vector was used for Southern analysis on a subset of clones. An *Nco*I digest hybridized to a 3′ external probe was used to ensure that the restriction pattern seen on the *Bam*HI and *Hin*dIII blots was due to targeting at the *hprt* locus (Fig. 1). An *Nco*I wild-type fragment is 23 kb long, and the insertion of IV6.8 will introduce an *Nco*I site (in the *neo*) to form a 21-kb fragment. Every clone analyzed with an *Nco*I digest had the 21-kb fragment. The combination of *Bam*HI and *Hin*dIII digests hybridized to the internal probes and the *Nco*I digest hybridized to the external probe verified the predicted insertion event at the *hprt* locus.

Two of nine TGʳ clones targeted with linearized RV6.8PGK had gene replacement events. A gene replacement event with RV6.8PGK should introduce *Bam*HI, *Hin*dIII, and *Nco*I sites into exon 3. To test the structure of the targeted locus, DNA from nine clones was digested with *Bam*HI and hybridized to exons 2 and 3 (Fig. 2). The wild-type 7.0-kb fragment should be cut into two fragments of 5.9 and 2.3 kb. Surprisingly, only two of nine TGʳ clones analyzed showed the predicted structure (Table 1), clones 3 and 4 (Fig. 2B). A *Hin*dIII digest and an *Nco*I digest hybridized to the external probe confirmed that only these clones were targeted by gene replacement. The *Hin*dIII digest shows that the wild-type 7.1-kb fragment is reduced to 6.6 kb, and the *Nco*I digest shows that the wild-type 23-kb fragment is reduced to 13 kb (Fig. 2B). As expected, these clones did not hybridize to plasmid sequences (data not shown).

Entire construct integrated in majority of targeted events generated with linearized RV6.8PGK. The *Bam*HI, *Hin*dIII, and *Nco*I digests used to identify the two replacement events also identified changes in seven targeted clones that could not be explained as gene replacement events (Fig. 2). Analysis of the *Bam*HI digest hybridized to exons 2 and 3 showed that clones 2 and 6 through 9 had targeted events that included the plasmid and made a 6.8-kb duplication of the homologous sequences, with the *neo* located in the 5′ duplicate, so that the 7.0-kb wild-type fragment was retained and new 4.8- and 5.9-kb fragments were formed. The 4.8-kb fragment hybridized to plasmid sequences (data not shown). The *Hin*dIII digest hybridized to the 3′ external probe showed only the 7.1-kb wild-type fragment. This indicates that the 3′ duplicate was unaltered. An *Nco*I-*Sal*I digest hybridized to the external probe identified the introduction of the *Sal*I site in the plasmid polylinker, which reduces the wild-type 23-kb fragment to 22 kb. The 22-kb fragment hybridized to plasmid sequences (data not shown). These restriction and hybridization patterns are consistent with integration of the entire vector rather than a replacement (Fig. 3A). Integration of the entire vector can occur via

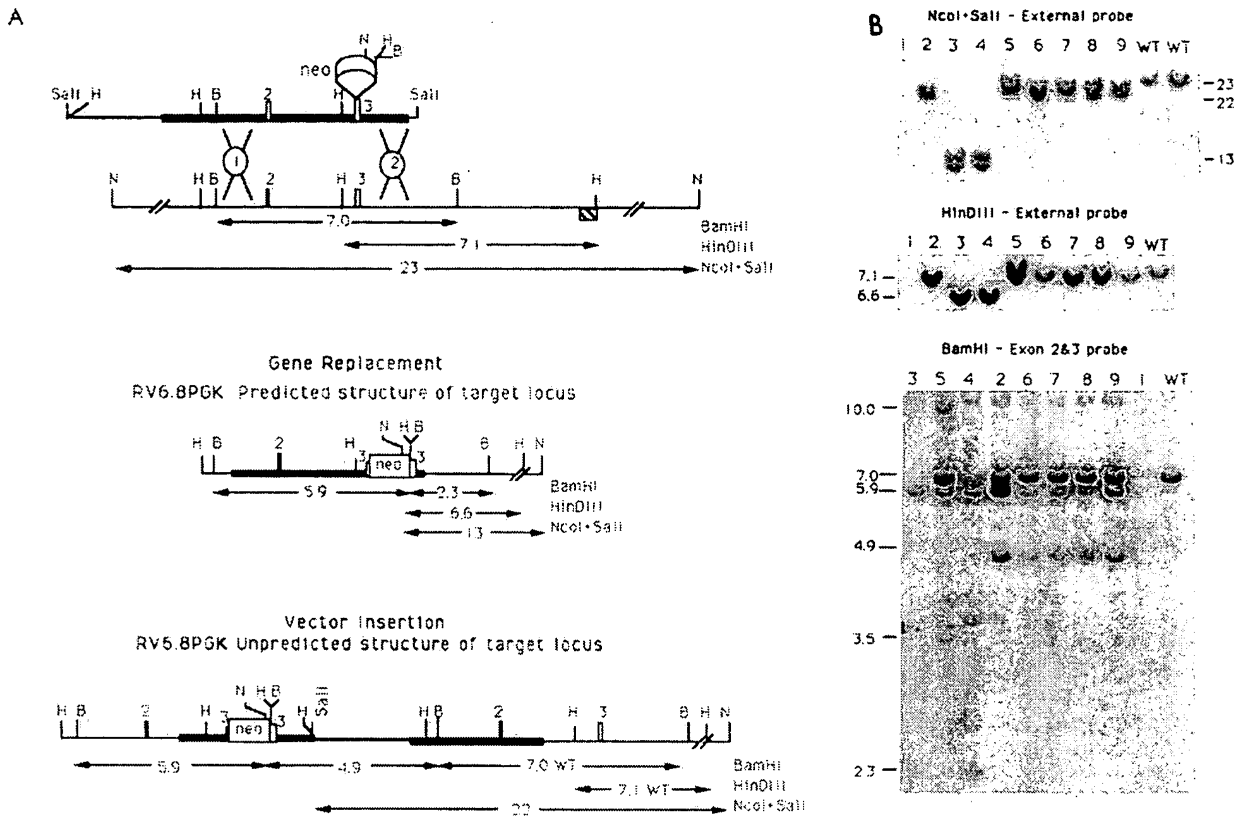


FIG. 2. Southern blot analysis of TG^r clones generated with RV6.8PGK. (A) Wild-type (WT) fragments and predicted sizes of fragments for double reciprocal recombination (gene replacement) and single reciprocal recombination (vector insertion) seen with a *Bam*HI (B) digest, a *Hin*dIII (H) digest, and an *Nco*I (N) digest. The thick line is *hprt* DNA of vector origin, the thin line is *hprt* DNA of genomic origin, and the line of intermediate thickness is plasmid DNA. PGKneobpA is a box labeled *neo*. Exons 2 and 3 are numbered rectangles and serve as the internal probe (taken from cDNA). The striped box is the external probe. The length of DNA is in kilobases. (B) The *Bam*HI digest was hybridized to exons 2 and 3. The *Hin*dIII and *Nco*I digests were hybridized to the external probe.

double reciprocal recombination of a concatemer on the long arm (the *neo* divides the homology into 5.6- and 1.2-kb fragments, the long and short arm) of the first unit, L1, and the long arm of a downstream unit designated unit x, Lx. This we refer to as an L1&Lx integration. The insertion may also occur via single reciprocal recombination on the long arm of a recircularized construct or circularized concatemer. This is referred to as vector insertion on the long arm (VILA).

Two TG^r clones targeted with RV6.8PGK gave restriction patterns that were not consistent with simple replacement or insertion events. In Fig. 2B, clone 5 appears to have been targeted by a modified insertion event. The *Hin*dIII and *Nco*I digests have the fragment sizes expected for an insertion event on the long arm. However, the *Bam*HI digest showed an aberrant pattern. Clone 1 has a large deletion that does not hybridize to exons 2 and 3 or to the external probe; this may be due to a spontaneous *hprt* deletion.

Majority of targeted clones generated with linearized RV6.8 occurred by an L1&Lx or VILA integration pattern. Ten TG^r clones were analyzed to determine the vector integration pattern of RV6.8 linearized with *Sal*I. A *Bam*HI digest hybridized to exons 2 and 3 and a *Hin*dIII digest hybridized to the external probe showed that the vector had inserted in eight clones (Table 1). The restriction pattern is similar to that seen for RV6.8PGK. Of the eight clones targeted by the integration of the entire construct, seven appeared to have integrated more than one unit. In Fig. 3C, clone 7 showed the insertion of a single unit and clone 8 shows the insertion of multiple units.

Clone 4 (Fig. 3C) was the result of a gene replacement event characterized by integration of a concatemer with one crossover on the long arm of the first unit (L1) and the other crossover on the short arm of a downstream unit designated unit x (Sx). We call this an L1&Sx integration pattern. By comparing the fragment densities of clone 4 with those of a single-unit insert, clone 7, it appears that unit x is the third or fourth unit in the concatemer. For the L1&Sx gene replacement event, fragments of 5.6, 4.2, and 2.3 kb are seen on a *Bam*HI digest hybridized to exons 2 and 3, and a fragment shift from the 7.1-kb wild-type fragment to 8.3 kb was seen with a *Hin*dIII digest hybridized to the external probe. One clone had a restriction pattern that is not representative of any of the possibilities discussed.

Southern analysis was also performed on seven clones targeted with RV6.8 cut with *Sac*I. Vector insertion was observed for four clones. All four clones appeared to have integrated multiple units. A modified integration was found in three clones (Table 1).

Majority of targeted clones generated with RV6.8 released from the plasmid occurred by integration of ligated units. To

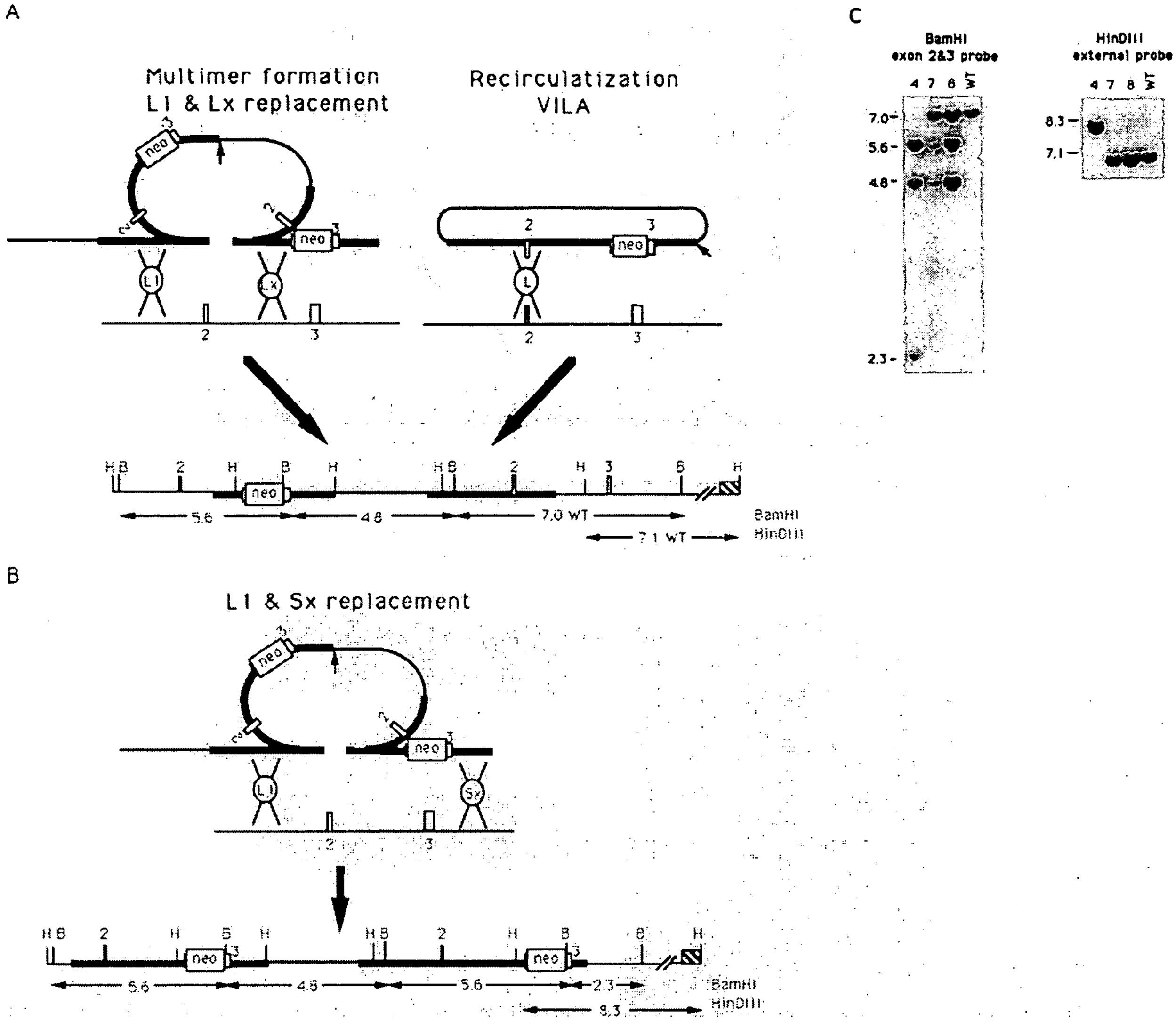


FIG. 3. Southern blot analysis of TG^r clones generated with RV6.8 linearized with *Sal*I. The thick line is *hprt* DNA of vector origin, the thin line is *hprt* DNA of genomic origin, and the line of intermediate thickness is plasmid DNA. MCIneopA is a box labeled *neo*. Exons 2 and 3 are numbered rectangles and serve as the internal probe (taken from cDNA). The striped box is the external probe. The length of DNA is in kilobases. The small arrows point to the location for end-to-end joining for multiple units of the construct. (A) Diagram of a multimer involved in gene replacement with the integration on the long arms, L1 and Lx. Diagram of a recircularized construct or circularized concatemer involved in VILA. Wild-type (WT) fragments and predicted sizes of fragments for gene replacement and vector insertion seen with a *Bam*HI (B) digest hybridized to exons 2 and 3 and a *Hin*dIII (H) digest hybridized to the external probe. (B) Diagram of a multimer involved in gene replacement with the crossover on the long arm of the first unit (L1) and the short arm of unit x (Sx). (C) The *Bam*HI digest was hybridized to exons 2 and 3. The *Hin*dIII digest was hybridized to the external probe. Sizes are in kilobases.

determine whether the proportion of vector replacement events would increase when the *hprt* homology was released from the plasmid, RV6.8 was cut with *Sac*I and *Sal*I and transfected into ES cells. The targeting frequency was unaltered (Table 1). Southern analysis was performed on TG^r clones with a *Bam*HI digest hybridized to exons 2 and 3 and a *Hin*dIII and *Nco*I digest hybridized to the external probe. We did not recover any gene replacement events in 15 clones analyzed. Instead, a variety of integration patterns that required the ligation of the *hprt* homologous sequence to itself and/or to the plasmid sequence prior to integration were seen (Fig. 4). The most common pattern appeared to occur by an L1&Lx or VILA integration of an *hprt*-plasmid-*hprt* concatemer, as seen in 6 of 15 clones analyzed with *Bam*HI, *Hin*dIII, and *Nco*I digests (Fig. 4B, clones 2, 5, 7, 8, 9, and 12). Clones 1, 3, and 6 are very similar to the forementioned samples in that the fragment shift on the *Nco*I digest is the same. However, a *Bam*HI digest hybridized to exons 2 and 3 revealed that although the 7.0- and 5.6-kb fragments were the same, the expected 4.8-kb fragment was only 4.6-kb. The size difference and hybridization pattern can be explained by limited exonuclease digestion associated with the end-to-end ligation (25). The pattern seen in clone 13 is an L1&Lx integration of an *hprt-hprt* dimer or a VILA

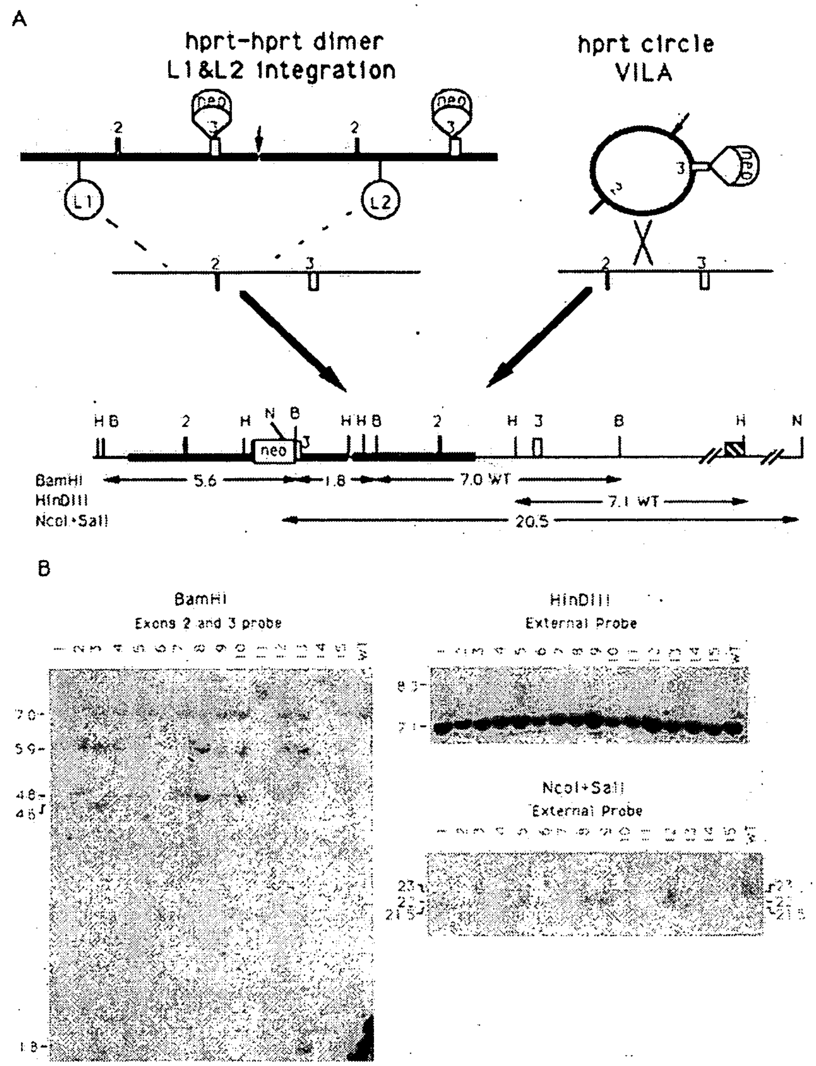


FIG. 4. Southern blot analysis of TG^r clones generated with RV6.8 cut with *Sac*I plus *Sal*I. The thick line is *hprt* DNA of vector origin, and the thin line is *hprt* DNA on genomic origin. MClneopA is a box labeled *neo*. Exons 2 and 3 are numbered rectangles and serve as the internal probe (taken from cDNA). The striped box is the external probe. The length of DNA is in kilobases. (A) Wild-type (WT) fragments and predicted sizes of fragments for gene replacement of an *hprt-hprt* dimer on the long arms of the first (L1) and second (L2) units or VILA of a recircularized *hprt* unit seen with a *Bam*HI (B) digest hybridized to exons 2 and 3, a *Hin*dIII (H) digest hybridized to the external probe, and an *Nco*I (N) digest hybridized to the external probe. The small arrows point to the location for end-to-end joining. (B) The *Bam*HI digest was hybridized to exons 2 and 3. The *Hin*dIII and *Nco*I digests were hybridized to the external probe. The 8.3-kb marker on the *Hin*dIII digest points to the fragment length seen for an L1&Sx integration.

integration of a circularized *hprt* unit (Fig. 4). The remaining samples were not readily interpreted with these digests and probes.

DISCUSSION

Insertion versus replacement vector: frequency. We have shown that an insertion vector targets *hprt* up to ninefold more frequently than a replacement vector with the same homologous sequences. Furthermore, since only 5% of the TG^r clones analyzed were L1&S1 replacement events, the ratio of absolute targeting frequencies for insertion events with an insertion vector (1.2×10^{-6}) to those for L1&S1 events with a replacement vector (1.3×10^{-8}) is 92:1. This observation is surprising and demonstrates that the configuration of the vector for the gene-targeting reaction can be rate limiting. The difference in the targeting frequencies implies that the vector-chromosome exchange is more favorable with an insertion vector. There are several aspects of our vectors that may have consequences for the rate of formation and/or resolution of recombination intermediates: crossover junctions (5), free ends, and uninterrupted homology.

(i) Crossover junctions. The formation and resolution of crossovers may be rate limiting. An insertion vector may require only a single point for crossover events, assuming that single reciprocal recombination occurs via the Meselson-Radding model (14), while a replacement vector may require two crossover points (22).

(ii) Free ends. Our data suggest that two adjacent homol-

ogous free ends are important in determining the frequency of gene targeting, contrary to another report (22). IV6.8 cut within the homology region to give two adjacent homologous free ends targets at a higher frequency than RV6.8 cut at either the 5′ or 3′ edge of homology to give one free homologous end or cut at both ends to give two nonadjacent homologous free ends. In yeast vectors, two adjacent homologous free ends are critically important in initiating the most efficient mechanism for recombination exchange, while vectors with one or no free homologous ends utilize a separate, less efficient mechanism (15). It is possible that ES cells also utilize two separate mechanisms for homologous recombination and that the most efficient mechanism requires two adjacent homologous free ends.

(iii) Uninterrupted homology. The replacement and insertion vectors are different in the configuration of homology in that the *neo* cassette interrupts the homology in the replacement vector but not in the insertion vector. It is possible that internal heterology might reduce the frequency of targeting for replacement vectors.

An earlier report has described the same targeting frequency for insertion and replacement vectors that target the 3′ end of *hprt* (22). The insertion vectors used in those studies are different from the one used here in that the 5′ and 3′ homologous ends were directly joined together and the *neo* gene interrupts the homology. In our vectors, the 5′ and 3′ *hprt* ends were joined to the plasmid and the *neo* marker, thereby retaining a long uninterrupted stretch of homology. Thus, the entire *hprt* homology is topologically more similar to the target gene; this may improve the gene-targeting efficiency.

Insertion versus replacement vector: mechanism. We have shown that our replacement vectors commonly integrated the entire construct, including the nonhomologous free ends, into the targeted locus. This was not expected because the *neo* gene was located within the homology and the vector was linearized at the edge of homology, so that gene replacement events should predominate (11, 22). There are two possible explanations for the integration of the entire replacement vector: gene replacement (double reciprocal recombination/gene conversion of a concatemer) and vector insertion (single reciprocal recombination of a recircularized vector or a circularized concatemer) (Fig. 3A).

If replacement vectors form a head-to-tail concatemer of two units, there are potentially four different patterns that would be identified on Southern blot, depending on the location of the crossover points. The *neo* gene divides the replacement vector homology into two parts, a 5′ long arm and a 3′ short arm. We have designated the first and second long arms as L1 and L2, respectively, and the first and second short arms as S1 and S2, respectively. Head-to-tail dimers could recombine on the long or short arm of the first and second unit to integrate the entire vector at the following crossover points: L1&L2, S1&S2, and L1&S2. Concatemers of greater than two units may integrate *x* units into the target locus, so that the crossover points may be L1&Lx, S1&Sx, and L1&Sx. A gene replacement event utilizing only one unit may also occur, L1&S1. Of the 17 clones identified to have integrated the entire construct, 13 appear to have integrated more than one unit when replacement vectors were linearized at a single site. Others have also observed concatemers at the target locus (18, 23, 24).

There are three possible recombination products if a single vector or a concatemer recircularizes. A single crossover point on either the long or short arm may occur and integrate the entire construct, or crossover junctions may form on both the long and short arm for a gene replacement event.

Southern analysis of targeted clones generated with replacement vectors showed that 23 of 41 targeted clones had insertion events that involved crossover points on only the long arm. This integration pattern cannot differentiate between the alternative pathways of a double crossover on L1 and Lx of a multimer or a single crossover on the long arm of a recircularized vector or circularized multimer. Targeting with a crossover point on the short arm may not be observed due simply to fivefold-lesser homology. An L1&Sx replacement was observed for only one clone. This integration pattern can only be generated by double reciprocal recombination of a multimer.

There have been two reports that showed that replacement vectors target *hprt* only by gene replacement (4, 22). However, neither identified targeted clones by Southern analysis with external probes, and both analyzed only one arm for targeting either by PCR (4) or by Southern analysis (22). The true structure of these targeted clones may not be fully understood. We show here that targeted colonies in which the entire vector is integrated will give wild-type fragments (*Hin*dIII digest, external probe for all L1&Lx and VILA integration patterns) or gene replacement fragments (*Hin*dIII digest, external probe for the L1&Sx integration), depending upon the digest and probes used. In order to fully understand the integration structure of targeted clones, the DNA must be tested with multiple digests and hybridized with probes both internal and external to the targeting sequences.

We released the plasmid from the targeting homology in a replacement vector in order to evaluate the effect it had on the targeting frequency and integration pattern. The targeting frequency was unaltered, and we did not detect any gene replacement events in 15 clones analyzed. Concatemers of *hprt*-plasmid-*hprt* were found in seven clones. Another three clones showed a very similar pattern, except that the 4.8-kb fragment appeared to be about 200 bp shorter, probably the result of exonuclease activity involved in the end-to-end ligation (25). One clone was shown to be an L1&L2 integration of an *hprt-hprt* dimer or VILA integration of a circularized *hprt* unit.

The inclusion of heterology on the free ends of a replacement vector in the recombination product has not been reported previously, perhaps owing to the selection and screening techniques used to isolate targeted events (2, 6, 7, 9, 11, 12, 20, 23, 27). Positive and negative selection involves the integration of a positively selectable marker that interrupts the homology and the removal of a negatively selectable marker located at the edge of homology (11). Targeted clones that incorporate external heterology will include an intact negatively selectable marker and die in selection. There has been one reported clone for which positive and negative selection has recovered a multiple repeat at the target locus although the negatively selectable markers were removed (23). We have observed clones targeted by L1&Lx or VILA of a replacement vector that show an alteration possibly due to exonuclease activity associated with the joining of free ends. Such an alteration may remove the negatively selectable marker and allow tandem repeats to survive in negative selection.

Use of PCR in screening for targeted clones involves amplification of a junction fragment that is only generated following homologous recombination (2, 7, 9, 12, 20, 27) and absolutely requires crossover events on the short arm of homology. The use of PCR will screen against integration

events with crossover points restricted to the long arm of the vector. Most of the TG^r clones targeted with RV6.8 and RV6.8PGK integrated the entire construct, with the crossover points restricted to the long arm; therefore, only 3 of 40 clones described here would be detected as positive by a PCR assay.

The insertion vector IV6.8 was shown to integrate the entire construct, as expected, although clones with gene replacement events would not integrate the *neo* and would die in selection. Our experiments do not determine whether the IV6.8 integration pattern was due to single reciprocal recombination at the adjacent homologous free ends or due to double reciprocal recombination of a concatemer, although the frequencies suggest that the former pathway is more likely. The differences in the targeting frequency between insertion and replacement vectors suggest that a different mechanism is occurring with insertion vectors, such as double-strand-break repair (21).

DNA transfected into a cell can clearly follow a variety of integration pathways. With replacement vectors, our data cannot discriminate whether end-to-end ligation to form concatemers or circles occurs more quickly than the targeted integration reaction or whether a concatenated or circularized molecule is a more suitable template for recombination. With insertion vectors electroporated at the same DNA concentration, we rarely observed insertions of concatemers. This suggests that the adjacent homologous free ends are rapidly assimilated in the targeting reaction and are not available for end-to-end joining. The differences in the targeting frequency between insertion and replacement vectors suggest that recircularized replacement vectors or circularized concatemers can use the more efficient pathway, although there are likely to be mechanistic differences in these events compared with those with vectors which have a double-strand break in homology.

ES cells appear to have a high affinity for targeting by the L1&Lx or VILA pathway and frequently incorporate heterologous free ends of the replacement vector into the target locus; the majority of events we analyzed were of this type. The use of vectors that utilize the most efficient pathway for homologous recombination may be important for genes that target at a low efficiency. The use of both PCR and negative selection with replacement vectors will enhance recovery of targeted clones and purify the population for replacement events; however, these strategies will also eliminate 95% of the targeted events. The observations of different classes of targeted recombinants has important implications for the design, screening, and analysis of targeted clones that will ultimately be used for genetic studies in transgenic mice.

ACKNOWLEDGMENTS

We acknowledge J. D. Wallace for technical assistance and thank our colleagues Mark Brenneman, Ann Davis, Ramiro Ramírez-Solis, John Wilson, and Hui Zheng for critical review of the manuscript.

This work was supported by grants from the NIH (A.B.), the Searle Scholars Program/Chicago Community trust (A.B.), the Granada Corporation (P.H.), and the Cystic Fibrosis Foundation (A.B., P.H., and J.R.).

REFERENCES

1. **Adra, C. N., P. H. Boer, and M. W. McBurney.** 1987. Cloning and expression of the mouse *pgk-1* gene and the nucleotide sequence of its promoter. Gene **60:**65–74.
2. **DeChiara, T. M., A. Efstratiadis, and E. J. Robertson.** 1990. A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting. Nature (London) **345:**78–80.
3. **Doetschman, T., R. G. Gregg, N. Maeda, M. L. Hooper, D. W. Melton, S. Thompson, and O. Smithies.** 1987. Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. Nature (London) **330:**576–578.
4. **Doetschman, T., N. Maeda, and O. Smithies.** 1988. Targeted mutation of the *Hprt* gene in mouse embryonic stem cells. Proc. Natl. Acad. Sci. USA **85:**8583–8587.
5. **Holliday, R.** 1964. A mechanism for gene conversion in fungi. Genet. Res. **5:**282–304.
6. **Johnson, R. S., M. Sheng, M. E. Greenberg, R. D. Kolodner, V. E. Palaioannou, and B. M. Spiegelman.** 1989. Targeting of nonexpressed genes in embryonic stem cells via homologous recombination. Science **245:**1234–1236.
7. **Joyner, A. L., W. C. Skarnes, and J. Rossant.** 1989. Production of a mutation in mouse *En-2* gene by homologous recombination in embryonic stem cells. Nature (London) **338:**153–156.
8. **Koller, B. H., P. Marrack, J. W. Kappler, and O. Smithies.** 1990. Normal development of mice deficient in β_2M class I proteins, and CD8$^+$ T cells. Science **248:**1227–1230.
9. **Koller, B. H., and O. Smithies.** 1989. Inactivating the β_2-microglobulin locus in mouse embryonic stem cells by homologous recombination. Proc. Natl. Acad. Sci. USA **86:**8932–8935.
10. **Kucherlapati, R. S., E. M. Eves, K. Song, B. S. Morse, and O. Smithies.** 1984. Homologous recombination between plasmids in mammalian cells can be enhanced by treatment of input DNA. Proc. Natl. Acad. Sci. USA **81:**3153–3157.
11. **Mansour, S. L., K. R. Thomas, and M. R. Capecchi.** 1988. Disruption of the proto-oncogene *int-2* in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. Nature (London) **336:**348–352.
12. **McMahon, A. P., and A. Bradley.** 1990. The *Wnt-1* (*int-1*) proto-oncogene is required for development of a large region of the mouse brain. Cell **62:**1073–1085.
13. **Melton, D. W., D. S. Konecki, J. Brennand, and C. T. Caskey.** 1984. Structure, expression, and mutation of the hypoxanthine phosphoribosyltransferase gene. Proc. Natl. Acad. Sci. USA **81:**2147–2151.
14. **Meselson, M. S., and C. M. Radding.** 1975. A general model for genetic recombination. Proc. Natl. Acad. Sci. USA **72:**358–361.
15. **Orr-Weaver, T. L., J. W. Szostak, and R. J. Rothstein.** 1981. Yeast transformation: a model system for the study of recombination. Proc. Natl. Acad. Sci. USA **78:**6354–6358.
16. **Pfarr, D. S., L. A. Rieser, R. P. Woychik, F. M. Rottman, M. Rosenberg, and M. E. Reff.** 1986. Differential effects of polyadenylation regions on gene expression in mammalian cells. DNA **5:**115–122.
17. **Schwartzberg, P. L., S. P. Goff, and E. J. Robertson.** 1989. Germ-line transmission of a c-*abl* mutation produced by targeted gene disruption in ES cells. Science **246:**799–803.
18. **Schwartzberg, P. L., E. J. Robertson, and S. P. Goff.** 1990. Targeted gene disruption of the endogenous c-*abl* locus by homologous recombination with DNA encoding a selectable fusion protein. Proc. Natl. Acad. Sci. USA **87:**3210–3214.
19. **Song, K., L. Chekuri, S. Rauth, S. Ehrlich, and R. Kucherlapati.** 1985. Effect of double-strand breaks on homologous recombination in mammalian cells and extracts. Mol. Cell. Biol. **5:**3331–3336.
20. **Soriano, P., C. Montgomery, R. Geske, and A. Bradley.** 1991. Targeted disruption of the c-*src* proto-oncogene leads to osteopetrosis in mice. Cell **64:**693–702.
21. **Szostak, J. W., T. L. Orr-Weaver, R. J. Rothstien, and F. Wong-Stahl.** 1983. The double-strand-break repair model for recombination. Cell **33:**25–35.
22. **Thomas, K. R., and M. R. Capecchi.** 1987. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell **51:**503–512.
23. **Thomas, K. R., and M. R. Capecchi.** 1990. Targeted disruption of the murine *int-1* proto-oncogene resulting in severe abnormalities in the midbrain and cerebellar development. Nature (London) **346:**847–850.
24. **Thompson, S., A. R. Clarke, A. M. Pow, M. L. Hooper, and D. W. Melton.** 1989. Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic

stem cells. Cell 56:313–321.
25. **Wilson, J. H., P. B. Berget, and J. M. Pipas.** 1982. Somatic cells efficiently join unrelated DNA segments end to end. Mol. Cell. Biol. 2:1258–1269.
26. **Wong, E. A., and M. R. Capecchi.** 1986. Analysis of homologous recombination in cultured mammalian cells in transient expression and stable transformation assays. Somat. Cell Mol. Genet. 12:63–72.
27. **Zijlstra, M., E. Li, F. Sajjadi, S. Subramani, and R. Jaenisch.** 1989. Germ-line transmission of a disrupted β_2-microglobulin gene produced by homologous recombination in embryonic stem cells. Nature (London) 342:435–438.

Exhibit 6

Int. J. Dev. Biol. 42: 943-950 (1998)

Thirteen years of manipulating the mouse genome: a personal history

ALLAN BRADLEY[1,2*], BINHAI ZHENG[1] and PENTAO LIU[1,3,#]

[1]Department of Molecular and Human Genetics, [2]Howard Hughes Medical Institute and [3]Program in Developmental Biology, Baylor College of Medicine, Houston, Texas, USA

ABSTRACT In 1974, Dr. Ralph Brinster published a paper describing the consequences of injecting embryonal carcinoma cells, the predecessors of embryonic stem cells, into mouse blastocysts. Despite their early promise, embryonal carcinoma cells would not efficiently populate the germ line of mice. A decade later Elizabeth Robertson and I described the efficient generation of germline chimaeras from cultured embryonic stem cells and shortly afterwards the genetic manipulation of the mouse germline using ES cells. Our demonstration of the potency of Embryonic Stem cells gave birth to a new era in manipulative mouse genetics, one in which endogenous genes can now be mutated at will using gene targeting of retroviral mutagenesis. This review focuses on the development and testing of concepts and techniques during the thirteen years after we knew germline modification of endogenous genes in the mouse would be possible. This period is one in which more and more sophisticated tools for manipulating the mouse germline were developed and implemented. In this review I have taken the rare opportunity to reveal some of my thought processes, frustrations, successes and failures as we moved through this exciting period of rapid technological change. As I look forward to the next thirteen years, I feel that this will be an equally exciting period for manipulative genetics as we struggle to formulate concepts and design experiments that enable us to understand gene function in an era when the sequence of all genes will be known.

KEY WORDS: *embryonic stem cells, gene targeting, hit and run, retroviral mutagenesis, chromosome engineering*

Germline transmission of ES cells

Fifteen years ago, the appearance of a pup with dark eyes in a litter caused great excitement in the Evans' laboratory. This pup was fathered by a male chimaera generated from cultured embryonic stem (ES) cells (Evans, and Kaufman, 1981, Fig. 1). Unbeknownst to us at the time, this germline transmission event signaled the emergence of a new age in mouse genetics. While we were confident that this experimental success was significant, the power of ES cells did not become apparent to the wider scientific community until a few years later when mutations were generated in ES cells in culture and transmitted into the mouse germline. Following the publication of our success in *Nature* (Bradley *et. al.*, 1984). I received a letter from Dr. Ralph Brinster. Unlike the thousands of letters I have received subsequently, this was not a request for materials but had a very simple message, "congratulations". I recall very clearly opening and reading this correspondence in the Tea Room of the University of Cambridge Genetics Department. I was honored that an individual of Dr. Brinster's stature had so selflessly taken the time to write a letter to a graduate student. Clearly, Dr. Brinster recognized the breakthrough.

The experiments which Elizabeth Robertson and I performed with ES cells were conceptually related to, though different in outcome from, experiments which Dr. Brinster had described ten

Abbreviations used in this paper: ES cells, Embryonic stem cells; EC cells, Embryonal carcinaoma cells; PCR, Polymerase chain reaction; Hprt, Hypoxanthine Phosphoribosyl transferase; LTR, Long terminal repeat; HSVtk, Herpes simplex virus thymidine kinase; FIAU, 1-(2'-Deoxy-2'-Fluoro-β-D-Arabinofuranosyl)-5- Iodouracil; LoxP, Locus of crossover P; Cre, Cyclization recombinase; HAT, Hypoxanthine, amniopterin, thymidine.

*Address for reprints: Howard Hughes Medical Institute, Baylor College of Medicine, Department of Molecular and Human Genetics, One Baylor Plaza, Houston, Texas 77030, USA. FAX: 713-798-8142. e-mail: abradley@bcm.tmc.edu

#Present address: ABL-Basic Research Program, Mammalian Genetics Laboratory, Building 539, 1214 Stillmeadow Place, #C, Frederick, MD 21702, USA

0214-6282/98/$10.00

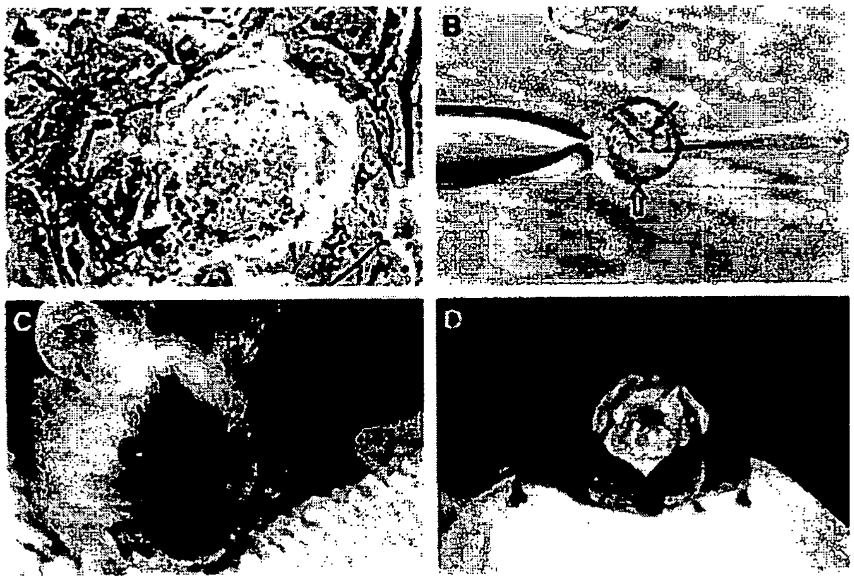


Fig. 1. Embryonic stem cell technology-reconstructing a mouse from cultured cells. **(A)** *ES cells growing on a STO feeder layer. White arrow shows a STO feeder cell nucleus, black arrow points to a colony of ES cells probably containing 1000 cells.* **(B)** *Injection of ES cells (thin arrow) into a 3.5 day blastocyst (thick arrow).* **(C)** *Chimaeric mouse showing contributions from descendants of the injected cells (pigmented) in the eye and melanocytes of the skin.* **(D)** *Germ line transmission from a chimaera demonstrated by the pigmented pups.*

years earlier. In this *Journal of Experimental Medicine* report (Brinster, 1974), Dr. Brinster showed that tumor derived embryonal carcinoma (EC) cells (the predecessors of ES cells) could be regulated by the embryonic environment of the blastocyst and contribute to the somatic tissues of chimaeric mice. This observation was subsequently confirmed by other laboratories (Mintz and Illmensee, 1975; Papaioannou *et al.*, 1975), however despite their early promise, EC cells were destined to diminish in popularity as an experimental system since the descendants of these cells rarely contributed to the germ line. Those successes that were reported could never be confirmed by other laboratories. The hope of being able to reconstruct the mouse germline from cultured EC cells therefore languished for almost ten years until it was eventually revived by our observation of germline transmission from chimeras constructed from cultured ES cells.

The demonstration of the germline transmission of ES cells took place during an exciting time for manipulative mouse genetics. The early-eighties was a time during which a series of landmark papers were published by the Brinster and Palmiter laboratories using transgenic mice generated by pronuclear injection (Brinster *et al.*, 1981,1983; Palmiter *et al.*, 1982). The power of gain of function modification of the mouse germ line as an experimental tool to address and answer important questions was made abundantly clear by these reports and this served as an inspiration to me. In 1984, ES cells offered a potential alternative route for generating transgenic mice, with the distinct advantage of being able to select or screen for a clone with a rare genetic change from millions of cells in culture *before* constructing a mouse. It was clear however that ES cells would not be able to compete with the efficiency with which transgenic mice could be generated by pro-nuclear microinjection, but we believed that we might be able to generate mutations in endogenous genes and thereby determine the function of these genes.

Today this dream has become reality. It is now virtually impossible to open a major journal without coming across one or more papers in which ES cells have been manipulated in culture to construct a loss of function mutation in a gene which is subsequently established in the germ line of mice. Like transgenics generated by pronuclear micro-injection, ES cell technology is now routinely practised in many laboratories to address a diverse array of biological questions. This article will focus on the evolution of ideas and approaches used to modify the genome of ES cells from a personal perspective, we apologize to those individuals whose work is not mentioned in this personal account.

The quest for recessive mutations

Shortly after we obtained germ line transmission of ES cells, we turned our attention to manipulating the genome of ES cells and establishing those mutations in the germ line. Two approaches were adopted, the first was retroviral mutagenesis using the helper free retroviral vectors, the second was homologous recombination.

Retroviral mutagenesis

Retroviral mutagenesis proved to be quite an efficient means of gene transfer in ES cells. However, the recombinant retroviral vectors available at the time had viral promoters which functioned very poorly in ES cells, so that the apparent viral titers on ES cells were greatly reduced compared with those which could be measured using fibroblasts. In order to achieve productive infection it was necessary to infect ES cell cultures with virus repeatedly. By following this strategy, adequate numbers of viral particles were eventually delivered to the culture so that every ES cell was infected multiple times, obviating the need for selection (Fig. 2). In the first instance, pools of ES cells from these experiments were injected into blastocysts to generate chimeras which transmitted and segregated these proviral insertions in the mouse germ line. These experiments provided the first evidence that genetically manipulated ES cells were still totipotent (Robertson *et al.*, 1986).

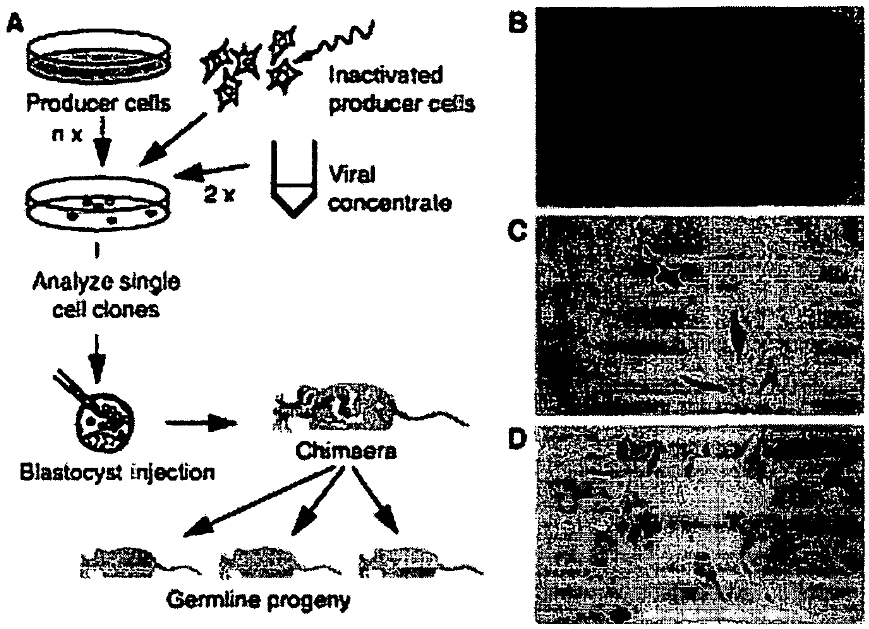


Fig. 2. Using retroviral mutagenesis to modify the mouse genome via ES cell Technology. (A) *The scheme used to demonstrate the germ line transmission of ES cells genomes modified in culture. Since retroviral promoters worked poorly in ES cells and active promoters which worked well in ES cells had not been identified, ES cells were infected multiple times with the virus or treated with a concentrated preparation of virions. The course of the infection could be monitored by Southern blots. Cloning of these cells could be accomplished by injecting the cells into blastocysts since only a few of the injected cells contributed to the germ line in any one chimaeric mouse.* **(B)** *F_1 Germ line progeny carried multiple copies of the provirus integration events distinguished by the use of enzymes which cut genomic DNA to generate unique junction fragments.* **(C and D)** *Primary fibroblasts from Hprt positive (C) and Hprt negative mice labeled with a ^{3}H-hypoxanthine incorporation assay confirming that the cells from the latter lack functional Hprt (D).*

Subsequently, some of these retroviral insertions were bred to homozygosity to identify integration events that were mutagenic and had inactivated interesting genes (Conlon *et al.*, 1991; Zhou *et al.*, 1993). Due to the random nature of retroviral insertions, only a subset of the integrations are actually mutagenic unless the insertions have been pre-selected in some way. Therefore in our experiments only about 5% of the integrations caused detectable phenotypes when bred to homozygosity.

These large pools of ES cells, carrying multiple independent retroviral insertions per cell, represent a vast collection of mutant cells which can be screened to identify integration events in specific genes (Fig. 2). In the first instance, we decided to recover mutations in a gene in which loss of function mutations could be directly selected in culture. *Hprt* is an X-linked gene, therefore it is present at only a single copy in XY ES cells and loss of function mutations can be directly selected in 6-thioguanine. *Hprt*-negative ES cell clones were recovered with this selection and these clones were used to generate chimeras which were bred to establish the mutant *Hprt* allele in the mouse germ line. Hprt deficient mice were subsequently generated by intercrossing F_1 animals (Kuehn *et al.*, 1987). These mice were the first to be generated with a specific modification of an endogenous gene through the modification of a cell line *in vitro*.

While retro-viral mutagenesis worked well for genes such as *Hprt*, where loss of function phenotypes could be selected in culture, it was hard to adopt this methodology for autosomal genes, where recessive mutations could not be directly selected. We proposed that we could identify clones with a viral insertion in a gene of interest by using a very deep library of ES cell clones each with multiple retroviral insertion events. By infecting at high multiplicity (100 insertions/clone) and generating a library of 10,000 clones theoretically the genome would be saturated with pro-viral insertions at an average density of 1 insertion every 3 kbp. In principle, one could screen such a library of ES cell clones using PCR primers specific for the gene of interest in combination with primers specific for the retroviral LTRs (Fig. 3). An appropriate junction fragment in the pool generated by an insertion in the locus of interest should be amplified with these primers. In principle, mutations in most genes would be accessible in this pool, provided the target locus was of a reasonable size. However, the work required to identify and recover specific clones from the pool is substantial, which is why this technique was not aggressively pursued once it was clear that mutations in endogenous genes could be efficiently generated by gene targeting techniques.

Many of the problems with the first generation of retroviral vectors have now been resolved. Specifically, integrations into genes can be directly selected by the use of read through transcription (von Melchner and Ruley, 1989) or by the use of splice acceptor and splice donor sequences that activate the expression of selectable markers (Friedrich and Soriano, 1991). Moreover, the integration loci are now readily identified which has meant that it is possible to generate vast libraries of ES cell clones each tagged with a unique insertion event (Hicks *et al.*, 1997; Zambrowicz *et al.*, 1998). My original view that retroviral mutagenesis would become the dominant technology for obtaining mutations seems likely to be correct, despite the current popularity of gene targeting. In part, this view has been realized by founding a commercial company to pursue this objective, Lexicon Genetics Inc., where the implementation of gene trapping on a genome wide scale was set as a primary goal.

Gene targeting-the early days

Simultaneously with our retroviral mutagenesis experiments, we began to work on trying to mutate genes by homologous recombination in ES cells. There was evidence in the literature that

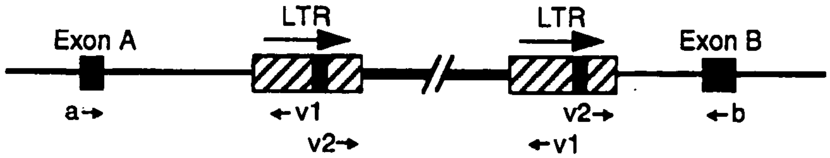


Fig. 3. A PCR screen used to identify retroviral insertion events in known genes in ES cells. *Pools of ES cells were screened for retroviral-exon junction fragments using primers specific to the LTR of the virus (v1 and v2) and primers specific to the exons of the target in the genome (a and b). Using infection ratios of 100 viral genomes per cell, we anticipated that this would give an effective coverage of 1 hit per 3 kb in a collection of 10^4 clones.*

mammalian cells could mediate extra-chromosomal homologous recombination between co-transfected DNA molecules (Kucherlapati *et al.*, 1984). However, there was no evidence that vector-chromosome recombination would be possible. The prevailing views at the time were that the mammalian genome was much too complex for incoming vector DNA to search, find and recombine with a homologous target before the efficient non-homologous recombination pathway effectively inserted the vector into a random location in the genome, how wrong this view was! A decade of gene targeting has revealed to us that homologous integration can often occur just as frequently as random insertion, moreover it is rare for the ratio of targeted to random insertions to be less than 5%.

In 1984, the absence of any published data that gene targeting was possible led me to take the view that strong positive selection for the desired recombination event was essential if our experiments were ever going to succeed. Our first experiments therefore attempted to target a gene which was highly expressed in ES cells *(c-myc)* and to use a selectable marker which lacked it's own promoter and would therefore only be expressed if it was inserted into the transcribed portion of the *c-myc* locus (Fig. 4). This vector failed to yield any targeted clones, to this day I don't understand why, especially since we later targeted *c-myc* using the same genomic clone at a 10% frequency (Davis *et al.*, 1993)!

Around the time my experiments to positively select for targeting at the *c-myc* locus was proving to be unsuccessful, there was emerging evidence that vector-chromosome recombination could occur in mammalian cells in two seminal papers published in 1985 from the Sternberg (Lin *et al.*, 1985) and Smithies (Smithies *et al.*, 1985) laboratories. In the former case, incoming vector DNA restored a defective *HSVtk* gene which had been randomly integrated into the genome while in the latter case, the endogenous β-globin locus was targeted with an insertion vector in human erythroleukemia cells. In the case of the β-globin targeting, the targeted clones could not be selected but had to be identified by an elegant yet very labor intensive sib-selection screen. These experiments suggested that targeted integration would occur at a frequency of 10^{-3} per random integration event. In the following year the Capecchi laboratory began to unravel some of the variables that affected gene targeting using a microinjection and positive selection approach in fibroblasts (Thomas *et al.*, 1986).

Gene targeting-disrupting nonselectable loci

In 1987, both the Capecchi and Smithies laboratories reported successful targeting of the *Hprt* locus in ES cells (Doetschman *et al.*, 1987; Thomas and Capecchi, 1987). These reports of gene targeting in ES cells pursued the same target we used to demonstrate the effectiveness of our retroviral mutagenesis experiments, the X-linked *Hprt* locus, relying on the direct selection for gene targeting events in 6-thioguanine (Thomas and Capecchi, 1987) or HAT (Doetschman *et al.*, 1987). Targeted clones were reported to be recovered at frequencies ranging from 10^{-3} to 10^{-5} per stably transfected cell. Moreover, Thomas *et al.* (1986), reported that efficient recombination seemed to require very large vectors, ideally 10 kb of homology (Thomas and Capecchi, 1987). We did not know this at the time, but these frequencies were several orders of magnitude lower than those we and others were subsequently able to achieve at many different loci.

At the time though, we viewed this data very positively since vector chromosome recombination had been achieved. However, the very low frequencies reported forced us to adopt a variety of elaborate selection and screening strategies for identifying recombinant clones. This included adapting vectors so that we could detect recombinant clones in pools by PCR amplification of junction fragments (McMahon and Bradley, 1990; Soriano *et al.*, 1991), the use of very large homology regions and the use of alternative positive selection schemes such as 3' trapping (Donehower *et al.*, 1992).

Approximately a year later the Cappecchi laboratory published a paper in *Nature* in which selection for targeting into non-selectable loci was achieved (Mansour *et al.*, 1988). This technique known as positive negative selection, heralded as a breakthrough, has become the most widely used technique in the

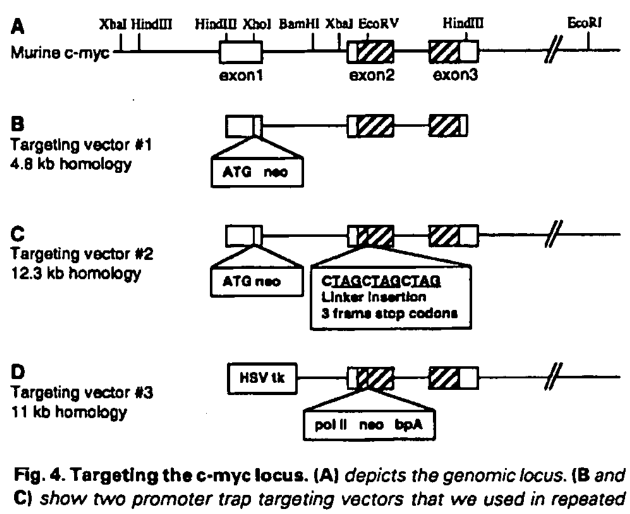


Fig. 4. Targeting the c-myc locus. (A) *depicts the genomic locus.* **(B and C)** *show two promoter trap targeting vectors that we used in repeated attempts to obtain targeting at the c-myc locus. These vectors failed to give any targeted clones while the vector in panel* **(D)** *targeted at a frequency of 9% of clones analyzed.*

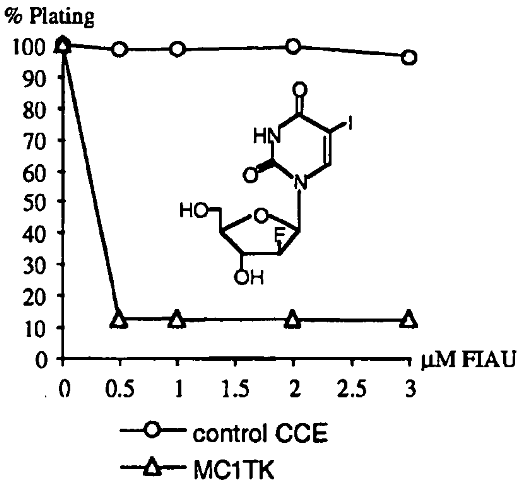


Fig. 5. Testing positive negative selection. *CCE ES cells were transfected with a construct which carried MC1neopA and HSVtk selection cassettes cloned into a plasmid backbone. Colony numbers were assessed after plating in G418 (positive) or G418+FIAU (positive/negative). The negative selection efficiency is calculated as the ratio of the colony number in the different selections (about an 8 fold enrichment) which was the best we ever achieved.*

practice of gene targeting. Ironically the power of the selection achieved with this technique is actually quite modest. In 1988, my two person laboratory tried very hard to repeat the experiments which had been published in *Nature* yet to this day neither we nor any other group I know, have ever achieved the 2,000 fold enrichment reported! The experiment illustrated in Figure 5 was repeated innumerable times under a vast array of varying experimental conditions [including with RV9.1TK, one of the vectors described in Mansour *et al.*, (1988)]. Initially, we believed that the problem lay in the efficiency of the selection. In 1988, the only readily available drug designed for HSV*tk* negative selection was acylovir which was very non-specific and killed wild type-cells almost as efficiently as HSV*tk* expressing cells. We took two approaches to attempt to resolve this problem. We explored the use of alterative Herpes Virues including marmoset and bovine herpes virus *thymidine kinase* genes which exhibited high activity and specificity for several nucleotide analogs which were readily available. Simultaneously, we obtained other analogs which were under development. The best and most specific enrichment we ever achieved was approximately 8 fold using an analog called FIAU! The failure to obtain better enrichments was not a failure of selection, (every FIAU-resistant clone that we have analyzed has lost or mutated the HSV*tk* gene), rather it appears that the enrichments are always quite modest.

Fortunately, the frequency of targeting at most loci (including *Hprt*) is many orders of magnitude higher than the frequencies first reported by Thomas and Capecchi (Thomas and Capecchi, 1987). Even without negative selection, it is not unusual to obtain a ratio of targeted to random insertion events of 5 to 20% in ES cells.

While we were realizing our goals using homologous recombination in ES cells, the Brinster and Palmiter laboratories had embarked on a heroic series of experiments to attempt to obtain productive recombination in zygotes. In these experiments 1841 mice were born including 506 transgenics, of which only one was targeted (Brinster *et al.*, 1989). In much the same way that our early attempts at targeting in ES cells were based on inadequate knowledge of the requirements for efficient recombination, the injected DNA used in these experiments was not a particularly good substrate for productive homologous recombination with the target locus. Subsequently we provided Dr. Brinster's laboratory with a construct that yielded a 10% targeting efficiency in ES cells, however this also proved to be ineffective in mediating recombination following zygote injection. I am frequently asked about the possibility of achieving targeting via pro-nuclear injection, I'm rarely enthusiastic! After all ES cells have proven to be the workhorse vehicle for modification of the mouse genome.

Developing new cell lines

In the process of obtaining ES cell clones with targeted mutations in many different loci, we began to observe dramatic differences in the behavior of these clones when they were used to generate chimaeric mice. Importantly, many of these subclones formed chimeras very inefficiently, and the chimeras that were generated were usually low grade and these never exhibited germ line transmission of their ES cell derived genome. Occasionally, we identified a clone from the same population which exhibited the high chimera forming efficiency characteristics exhibited by its parental population. Other laboratories with whom we were competing on a variety of different projects also ran into the same technical difficulties.

I realized that we were observing the problem of genetic drift in our parental ES cell lines. Abnormal cells were cloned out in the

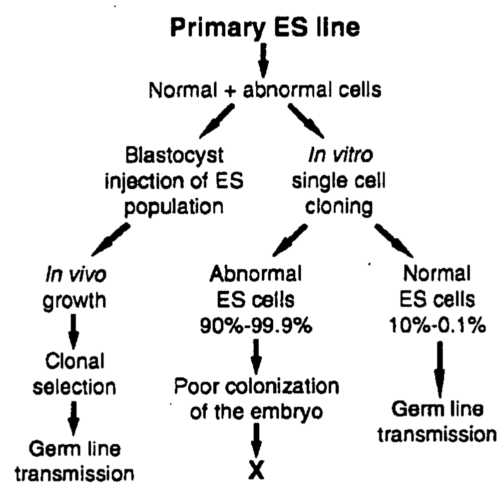


Fig. 6. Clonal variation in ES cell lines. *During the growth of ES cells variants will arise which will overgrow the normal population. These variants will not normally compromise the ability of the parental pool to contribute to the germ line, since usually adequate numbers of "normal" cells are still present in the population. When the parental cells are subcloned, the "abnormal" subclones will be identified, these will not contribute to chimeras or the germ line efficiently. These "abnormal" cells may represent the majority of subclones analyzed. Subclones of normal ES cells will have a much greater efficiency in forming chimeras and germ line transmission than the parental population.*

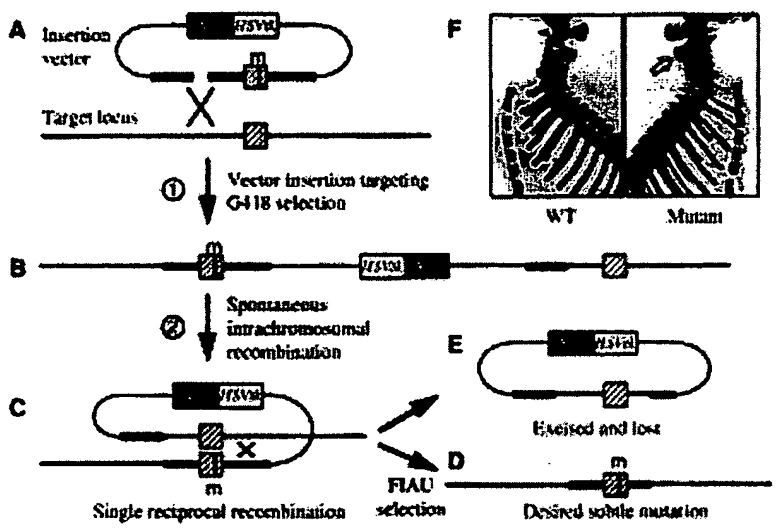


Fig. 7. The Hit and Run targeting scheme which we developed to generate point mutations in the genome (Hasty *et al.*, 1991). (A) *An insertion targeting vector was configured for both positive (neo) and negative (HSVtk) selection. This vector was modified to carry the desired point mutation and was linearized prior to transfection.* **(B)** *Recombinant clones were identified by Southern analysis.* **(C)** *The duplicated sequences in the recombinant locus spontaneously recombine to pop-out the vector either by an intra-chromosomal event (shown) or a sister chromatid pathway (not shown).* **(D)** *The revertant cells are selected for the loss of HSVtk in FIAU and screened for the desired allele.* **(E)** *The excised circle is lost during normal cell replication.* **(F)** *Skeletal preparations of the first mutant mice generated with a point mutation via ES cell manipulation. The embryo on the left is wild type, while the one on the right has a homeotic transformation in the Axis which has been converted to the Atlas (arrow) due to a premature termination in the third helix of the homeobox of HoxB4 (Ramirez-Solis et al., 1993).*

process of selecting and screening for targeted clones, and these were never able to colonize the germ line. The injection of the original populations of these ES cells into blastocysts indicated that normal cells were still present in these cultures, and that these normal cells could still contribute to chimeras even in the presence of large numbers of abnormal ES cells (Fig. 6). Our early recognition of this problem led me to isolate and evaluate new clones of STO feeder cells, one of which (SNL76/7) had dramatically improved qualities for ES cell maintenance compared with the parental cell line. Using these cells as feeder layers three new ES cell lines were isolated (AB1, AB2.1 and AB2.2) which we have used almost exclusively for the last 10 years (McMahon and Bradley, 1990; Soriano *et al.*, 1991; Donehower *et al.*, 1992; Matzuk *et al.*, 1992; Ramirez-Solis *et al.*, 1995). We still obtain very high germline transmission rates with the targeted clones derived from these cell lines. Our chromosome engineering projects require 3 successive subcloning steps on these cell lines, however, this has never proved to be a problem for obtaining germline transmission of these modified chromosomes (Ramirez-Solis *et al.*, 1995).

These new cell lines gave my then very small laboratory a competitive edge in efficiently transmitting various mutations into the mouse germline. When some of the early successes with these cells were published this created an unusual challenge for my small group to distribute these cells to the 1,000-plus laboratories that have now requested them. Distributing materials is usually a pretty thankless task, occasionally one receives a letter of acknowledgment, rarely anything more. My group remembers the thanks we received for dispatching cells to Ralph Brinster, gourmet cheese and French red wine, for once I felt we got the best part of the exchange!

Generating point mutations in the mouse genome

As we began to realize that gene targeting frequencies were actually exceedingly high, we shifted our attention to generating subtle mutations in endogenous genes. One of my concerns at the time related to our ability to interpret mutations we were generating in the *HoxB4* gene, one of the dense cluster of genes in the *HoxB* cluster on mouse chromosome 11. My concern related to the possible effect that the insertion of a selection cassette with strong promoters and enhancers in the *HoxB4* gene might have on the regulation of other genes in the region. To obviate this concern we explored three methods for making point mutations in the genome. The first of these involved co-electroporating two cassettes in to ES cells. One cassette served to facilitate selection for the transfected cells while the second was designed to recombine with the target locus. Although we were able to achieve the desired outcome with this strategy, the propensity of co-transfected DNA to co-integrate at a single locus made this technique very inefficient (Davis *et al.*, 1992). The second technique we explored was to use two rounds of targeting with replacement vectors. This required an initial recombination event to insert a negatively selectable maker into the locus and a second round of recombination to replace the cassette in the target locus with one that contains the desired modification. This technique has since been published by other laboratories (Askew *et al.*, 1993; Stacey *et al.*, 1994) and called tag-and exchange or double replacement gene targeting, respectively. While we could also occasionally obtain the desired recombinations with this technique, I never felt that this method was robust enough to warrant publication.

By contrast, a technique we termed "Hit and Run" proved to be highly effective (Hasty *et al.*, 1991). Simultaneously, the Smithies laboratory reported the same concept which they termed "In and Out" (Valancius and Smithies, 1991). This technique relies on a two step recombination procedure (Fig. 7). In the first step an insertion vector was constructed with the desired point mutation in the homologous sequences. In the case of *HoxB4* this was in the third helix of the homeodomain and was predicted to give a *null* allele. This insertion vector contained both positively and negatively selectable elements in the vector backbone. This insertion

vector was used to recover recombinants in the first step, utilizing the positive selectable marker. In the second step these recombinants were selected for "pop-out" events due to the intrachromosomal or sister chromatid recombination event between the duplication generated by the insertional targeting. These events were selected using the negative selection cassette in the vector to eliminate non-popped out clones. This method proved to be highly efficient and we generated the desired mutation in the *HoxB4* locus. The first mice with an engineered part mutation were described by my group in 1993. As we predicted, the mice exhibited a different phenotype compared with an allele in the same locus in which we had inserted a selectable marker(Ramirez-Solis *et al.*, 1993). We have generated other mutations in the *HoxB* cluster using this technique (Studer *et al.*, 1996). Other groups have since recognized the problem of the long range consequences of the insertion of selectable markers on neighboring transcription units (Olson *et al.*, 1996). These effects can extend over distances greater than 10 kb.

Long range recombination

In 1992, I decided to move the laboratory in a new direction. As far as I was concerned, most of the outstanding issues required for effective gene targeting had been solved and we could make most mutations we desired. At the same time, I saw an opportunity to engineer the mouse so that we could effectively use it in genetic screens. In particular, I had the desire to perform screens in a haploid context, recognizing the power of deficiency screens in *Drosophila* (Bridges, 1917) and the power of haploid genetics in general. However, a major obstacle lay in the way, namely the ability to generate deletions in a directed way in ES cells. We attempted to make large deletions in a single step with the use of standard replacement vectors. Although we succeeded in making 19 kb deletions at the *Hprt* locus (Zhang *et al.*, 1994), it was never possible for us to obtain larger deletions at an autosomal locus with this technique, even with the use of very powerful selection schemes analogous to those used at the *Hprt* locus. Therefore a different technique to generate and select clones with the desired deletions needed to be developed.

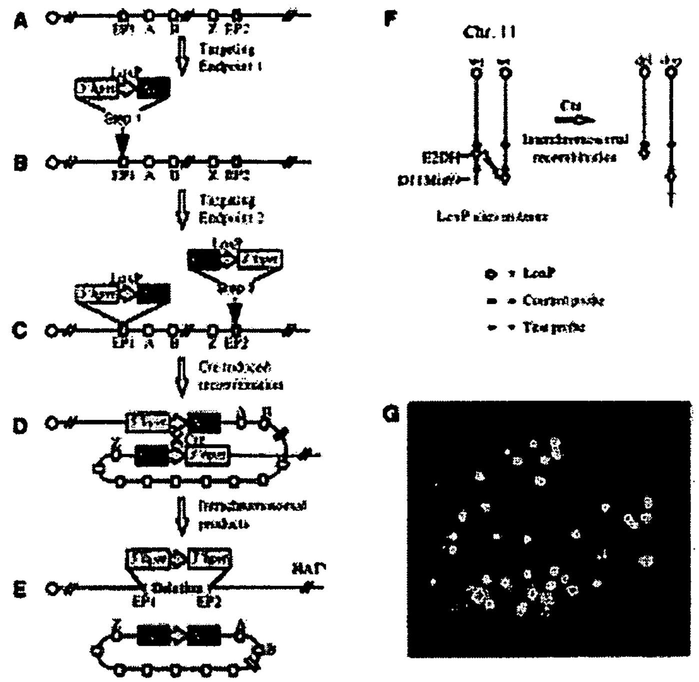


Fig. 8. Chromosome engineering in ES cells (Ramirez-Solis *et al.*, 1995). (A) *The unmodified chromosome containing all the genes between the deletion endpoint 1 (EP1) and the endpoint 2 (EP2) genes A-Z.* **(B)** *Targeting the loxP-deletion vector to EP1 using G418 selection and screening for targeted clones.* **(C)** *Targeting the second loxP-deletion endpoint vector to EP2 in cells carrying the first targeting event, using puromycin selection (only targeting in cis is shown for simplicity).* **(D)** *Cre expression in cells with the two endpoints targeted mediates recombination between the loxP sites at the deletion endpoints.* **(E)** *Cre recombination results in a deletion chromosome which is selected in HAT because the recombination reconstructs the Hprt gene from two non-functional parts and a ring chromosome which is lost during the normal process of cell division.* **(F)** *Cre-mediated recombination between deletion endpoints targeted to the two homologs also occurs. The recombination product is also selectable in HAT and is the deletion chromosome. However, the reciprocal product is not lost as a chromosome ring but is retained as a duplication on the homologous chromosome.* **(G)** *FISH confirms the generation of a duplication and a deletion chromosome from the trans recombination event.*

We elected to attempt to generate deletions using the *loxP*-Cre site specific recombinase system. Assuming that long range *loxP*-*Cre* recombination would probably be very inefficient a positive selection scheme was designed that would enable us to recover clones with the desired modification event. To do this, we divided an *Hprt* mini gene cassette into two overlapping but non-functional components where each piece included a *loxP* site. To generate deletions we had to construct a "pre-deletion chromosome" in which the cassettes with the *loxP* sites were targeted to the appropriate positions (Fig. 8). This was somewhat of a risk since we would not know if Cre would function over such large distances until the final step of the experiment namely, the expression of the Cre recombinase in the double targeted cell and the recovery of HAT resistant clones. In a series of experiments, we sequentially targeted the deletion endpoints on mouse chromosome 11 and tested the concept. Remarkably, we found that *Cre* recombinase would function over vast distances in the mouse genome at very high efficiencies. We could generate all classes of genomic rearrangements, deletions, duplications and inversions; chromosome engineering was born. (Ramirez-Solis *et al.*, 1995).

We have subsequently used this technology to engineer many different re-arrangements on mouse chromosome 11, including very large rearrangements of up to 70% of the chromosome, including balancer chromosomes tagged with recessive lethal mutations and coat color markers. There are many projects in the laboratory which are using these genetic reagents in screens with the goal of building a functional map of genetic elements from this region of the mouse genome.

Conclusion

The specificity with which either a gene or chromosomal region can be modified and the range of alterations that are now possible using the genome of ES cells as a surrogate mouse has made gene

targeting the dominant technology for manipulating the mouse genome for the last eight years. Making mutations by gene targeting is time consuming and for many laboratories rate limiting. In the genomics era it is desirable to obtain mutations without a customized approach for each gene. By solving the problem of identifying the integration site of the virus and implementing high throughput strategies, retroviral mutagenesis has re-emerged as a viable technology and is likely to establish itself as the dominant technology over the next few years. Homologous recombination will still have a place in the genomics era. Although *null* alleles generated by high throughput approaches are a good place to start the analysis of gene function, more detailed and specific questions based on knowledge of the gene product and gene structure are usually desirable and these can only be generated through customized approaches.

Acknowledgments

Work in the authors' laboratory has been supported over the last 13 years by the Searle Scholars Trust, The leukemia Society, The National Institutes of Health and the Howard Hughes Medical Institute.

References

ASKEW, G.R., DOETSCHMAN, T. and LINGREL, J.B. (1993). Site-directed point mutations in embryonic stem cells: A gene-targeting tag-and-exchange strategy. *Mol. Cell. Biol. 13(7):* 4115-4124.

BRADLEY, A., EVANS, M.J., KAUFMAN, M.H. and ROBERTSON, E.J. (1984). Formation of germ line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature 309:* 255-256.

BRIDGES, C. (1917). Deficiency. *Genetics 2:* 445-465.

BRINSTER, R.L. (1974). The effect of cells transferred into the mouse blastocyst on subsequent development. *J. Exp. Med. 140:* 1049-1056.

BRINSTER, R.L., BRAUN, R.E., LO, D., AVARBOCK, R., ORAM, F. and PALMITER, R.D. (1989). Targeted correction of a major histocompatibility class II Ea gene by DNA microinjected into mouse eggs. *Proc. Natl. Acad. Sci. USA 86:* 7087-7091.

BRINSTER, R.L., CHEN, H.Y., TRUMBAUER, M., SENEAR, A.W., WARREN, R. and PALMITER, R.D. (1981). Somatic expression of herpes thymidine kinase in mice following injection of a fusion gene into eggs. *Cell 27:* 223-231.

BRINSTER, R.L., RITCHIE, K.A., HAMMER, R.E., O'BRIEN, R.L., ARP, B. and STORB, U. (1983). Expression of a microinyected immunolhobulin gene in the spleen of transgenic mice. *Nature 306:* 332-336.

CONLON, F.L., BARTH, K.S. and ROBERTSON, E.J. (1991). A novel retrovirally induced embryonic lethal mutation in the mouse: assessment of the development fate of embryonic stem cells homozygous for the 413.d proviral integration. *Development 111(4):* 969-981.

DAVIS, A.C., WIMS, M. and BRADLEY, A. (1992). Investigation of coelectroporation as a method for introducing small mutations into embryonic stem cells. *Mol.Cell. Biol. 12(6):* 2769-2776.

DAVIS, A.C., WIMS, M. and BRADLEY, A. (1993). A null c-myc mutation caused lethality before 10.5 days of gestation in homozygotes and reduced fertioity in heterozygous female mice. *Genes Dev. 7:* 671-682.

DOETSCHMAN, T., GREGG, G., MAEDA, N., HOOPER, M.L., MELTON, D.W., THOMPSON, S. and SMITHIES, O. (1987). Targeted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature 330:* 576-578.

DONEHOWER, L.A., HARVEY, M., SLAGLE, B.L., MCARTHUR, M.J., MONTGOMERY, C.A.J., BUTEL, J.S. and BRADLEY, A. (1992). Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature 356(6366):* 215-221.

EVANS, M.J. and KAUFMAN, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature 292:* 154-156.

FRIEDRICH, G. and SORIANO, P. (1991). Promoter traps in embryonic stem cells: A genetic screen to identify and mutate developmental genes in mice. *Genes Dev. 5:* 1513-1523.

HASTY, P., RAMIREZ-SOLIS, R., KRUMLAUF, R. and BRADLEY, A. (1991). Introduction of a subtle mutation into the *Hox-2.6* locus in embryonic stem cells. *Nature 350(6315):* 243-246.

HICKS, G.G., SHI, E.-G., LI, X.-M., PAWLAK, M. and RULEY, H.E. (1997). Functional genomics in mice by tagged sequence mutagenesis. *Nature Genet. 16:* 338-344.

KUCHERLAPATI, R.S., EVES, E.M., SONG, K., MORSE, B.S. and SMITHIES, O. (1984). Homologous recombination between plasmids in mammalian cells can be enhanced by treatment of input DNA. *Proc. Natl. Acad. Sci. USA 81:* 3153-3157.

KUEHN, M.R., BRADLEY, A., ROBERTSON, E.J. and EVANS, M.J. (1987). A potential animal model for Lesch-Nyhan syndrome through introduction of HPRT mutations into mice. *Nature 326:* 295-298.

LIN, F.-L., SPERLE, K. and STERNBERG, N. (1985). Recombination in mouse L cells between DNA introduced into cells and homologous chromosomal sequences. *Proc. Natl. Acad. Sci. USA 82:* 1391-1395.

MANSOUR, S.L., THOMAS, K.R. and CAPECCHI, M.R. (1988). Disruption of the protocogene int-2 in mouse embryo-derived stem cells. A general stategy for targeting mutations to non-selectable genes. *Nature 336:* 348-352.

MATZUK, M.M., FINEGOLD, M.J., SU, J.-G.J., HSUEH, A.J.W. and BRADLEY, A. (1992). Alpha-inhibin is a tumour-suppressor gene with gonadal specificity in mice. *Nature 360:* 313-319.

MCMAHON, A.P. and BRADLEY, A. (1990). The Wnt-1 (Int-1) protooncogene is required for the development of a large region of the mouse brain. *Cell 62:* 1073-1085.

MINTZ, B. AND ILLMENSEE, K. (1975). Normal genetically mosaic mice produced from malignant teratocarcinoma cells. *Proc. Natl. Acad. Sci. USA 75:* 6247-6251.

OLSON, E.N., ARNOLD, H.-H., RIGBY, P.W.J. and WOLD, B.J. (1996). Know your neighbors: Three phenotypes of null mutants of the myogenic bHLH gene *MRF4*. *Cell 85:* 1-4.

PALMITER, R.D., BRINSTER, R.L., HAMMER, R.E., TRUMBAUER, M.E., ROSENFELD, M.G., BIRNBERG, N.C. and EVANS, R.M. (1982). Dramatic growth of mice that develop from eggs microinjected with metallothionein - growth hormone fusion genes. *Nature 300:* 611-615.

PAPAIOANNOU, V.E., MCBURNEY, M.W., GARDNER, R.L. and EVANS, M.J. (1975). Fate of teratocarcinoma cells injected into early mouse embryos. *Nature 258:* 70-73.

RAMIREZ-SOLIS, R., LIU, P. and BRADLEY, A. (1995). Chromosome engineering in mice. *Nature 378:* 720-724.

RAMIREZ-SOLIS, R., ZHENG, H., WHITING, J., KRUMLAUF, R. and BRADLEY, A. (1993). *Hox-b4* (*Hox-2.6*) mutant mice show homeotic transformation of a cervical vertebra and defects in the closure of the sternal rudiments. *Cell 73:* 279-294.

ROBERTSON, E., BRADLEY, A., KUEHN, M. and EVANS, M. (1986). Germline transmission of genes introduced into cultured pluripotent cells by retroviral vector. *Nature 323:* 445-448.

SMITHIES, O., GREGG, R.G., BOGGS, S.S., KORALEWSKI, M.A. and KUCHERLAPATI, R.S. (1985). Insertion of DNA sequences into the human chromosomal β-globin locus by homologous recombination. *Nature 317:* 230-234.

SORIANO, P., MONTGOMERY, C., GESKE, R. and BRADLEY, A. (1991). Targeted disruption of the *c-src* proto-oncogene leads to osteopetrosis in mice. *Cell 64:* 693-702.

STACEY, A., SCHNIEKE, A., MCWHIR, J., COOPER, J., COLMAN, A. and MELTON, D.W. (1994). Use of double-replacement gene targeting to replace the murine α-lactalbumin gene with its human counterpart in embryonic stem cells and mice. *Mol. Cell. Biol. 14(2):* 1009-1016.

STUDER, M., LUMSDEN, A., ARIZA-MCNAUGHTON, L., BRADLEY, A. and KRUMLAUF, R. (1996). Altered segmental identity and abnormal migration of motor neurons in mice lacking Hoxb-1. *Nature 384(6610):* 630-634.

THOMAS, K.R. and CAPECCHI, M.R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell 51:* 503-512.

THOMAS, K.R., FOLGER, K.R. and CAPECCHI, M.R. (1986). High frequency targeting of genes to specific sites in the mammalian genome. *Cell 44:* 419-428.

VALANCIUS, V. and SMITHIES, O. (1991). Testing an "In-Out" targeting procedure for making subtle genomic modifications in mouse embryonic stem cells. *Mol. Cell. Biol. 11:* 1402-1408.

VON MELCHNER, M. and RULEY, E. (1989). Identification of cellular promoters by using a retrovirus promoter trap. *J. Virol. 63:* 3227-3233.

ZAMBROWICZ, B., FRIEDRICH, G., BUXTON, E., LILLEBERG, S., PERSON, C. and SANDS, A. (1998). Disruption and sequence identification of 2,000 genes in mouse embrynoic stem cells. *Nature 392(6676):* 608-611.

ZHANG, H., HASTY, P. and BRADLEY, A. (1994). Targeting frequency for deletion vectors in embryonic stem cells. *Mol. Cell. Biol. 14:* 2404-2410.

ZHOU, X., SASAKI, H., LOWE, L., HOGAN, B.L. and KUEHN, M.R. (1993). Nodal is a novel TGF-beta-like gene expressed in the mouse node during gastrulation. *Nature 361(6412):* 543-547.

Exhibit 7

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 15/27, A01K 67/027, C12Q 1/68, C07K 15/06, C07G 15/00, A61K 37/02

A1

(11) International Publication Number: **WO 95/23862**

(43) International Publication Date: 8 September 1995 (08.09.95)

(21) International Application Number: PCT/AU94/00103

(22) International Filing Date: 4 March 1994 (04.03.94)

(71) Applicant *(for all designated States except US)*: LUDWIG INSTITUTE FOR CANCER RESEARCH [AU/AU]; Melbourne Tumour Biology Branch, Royal Melbourne Hospital, Royal Parade, Parkville, VIC 3050 (AU).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: DUNN, Ashley, Roger [GB/AU]; Ludwig Institute for Cancer Research, Melbourne Tumour Biology Branch, The Royal Melbourne Hospital, Royal Parade, Parkville, VIC 3050 (AU). STANLEY, Edouard, Guy [AU/GB]; National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA (GB). LIESCHKE, Graham, John [AU/AU]; Ludwig Institute for Cancer Research, Melbourne Tumour Biology Branch, The Royal Melbourne Hospital, Royal Parade, Parkville, VIC 3050 (AU). GRAIL, Dianne [AU/AU]; Ludwig Institute for Cancer Research, Melbourne Tumour Biology Branch, The Royal Melbourne Hospital, Royal Parade, Parkville, VIC 3050 (AU). FOWLER, Kerry, J. [AU/AU]; Ludwig Institute for Cancer Research, Melbourne Tumour Biology Branch, The Royal Melbourne Hospital, Royal Parade, Parkville, VIC 3050 (AU).

(74) Agent: SANTER, Vivien; Griffith Hack & Co., 509 St Kilda Road, Melbourne, VIC 3004 (AU).

(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: ANIMALS WITH TARGETED GENE DISRUPTION

(57) Abstract

The invention provides transgenic non-human animals in which genes encoding colony-stimulating factors have been specifically disrupted, so that the animals do not express the respective factors. In specific embodiments the invention provides mice in which expression of GM-CSF, G-CSF and/or CSF-1 is disrupted. These animals do not produce any bioactive CSF of the respective type. They are useful as models for a variety of conditions, and particularly for testing of therapeutic methods.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgystan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SI	Slovenia
CI	Côte d'Ivoire	KZ	Kazakhstan	SK	Slovakia
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DE	Germany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				

ANIMALS WITH TARGETED GENE DISRUPTION

This invention relates to animals having specifically designed alterations in pre-existing endogenous genetic loci, eg. animals having targeted disruptions of specific genes. In particular, the invention relates to animals in which the gene for a colony-stimulating factor is disrupted so as to create a null allele, whereby the animal does not express any detectable colony-stimulating factor. The colony-stimulating factor is granulocyte-macrophage colony-stimulating factor or granulocyte colony-stimulating factor. In a preferred embodiment of the invention, the animal is a rodent such as a mouse.

BACKGROUND OF THE INVENTION

Granulocyte-macrophage colony-stimulating factor (GM-CSF) is a haematopoietic growth factor which *in vitro* stimulates the survival, proliferation, differentiation and function of myeloid cells and their precursors, particularly neutrophil and eosinophil granulocytes and monocyte/macrophages (for review, see 1). The *in vivo* effects of GM-CSF have been studied in murine models by injecting pharmacological doses of GM-CSF (2), by generating GM-CSF transgenic mice (3), and by reconstituting lethally irradiated mice with bone marrow cells overproducing GM-CSF (4).

These studies confirm the haematopoietic activity of GM-CSF *in vivo*, and suggest that excess levels of GM-CSF may be implicated in some disease processes. However, the usual physiological role of GM-CSF is not well defined (5). Endogenously-produced GM-CSF is not usually detectable in serum (6), and in humans altered serum levels have not correlated clearly with haematological or disease processes (1,6). It has been suggested that GM-CSF may be produced and act locally (1), but the cells producing GM-CSF *in vivo* have yet to be identified. Moreover, it is not clear

whether GM-CSF is an essential regulator for steady-state production of granulocytes and macrophages, or whether it is required as a regulator for emergency haematopoiesis in response to challenges such as bacterial infection. It is also not known whether GM-CSF is involved in the normal development of non-haematopoietic tissues.

Granulocyte colony-stimulating factor (G-CSF) is a haematopoietic growth factor which *in vitro* controls granulopoiesis. It stimulates the survival, proliferation, differentiation and function primarily of neutrophil granulocytes. As in the case of GM-CSF, the *in vivo* effects of G-CSF have been studied in murine models by injection of pharmacological doses of G-CSF, and by reconstitution of lethally irradiated mice with bone marrow cells transformed with a retroviral vector carrying cDNA encoding G-CSF (reviewed in Reference 7). However, as with GM-CSF, the usual physiological role *in vivo* of G-CSF is unclear. It may act as a regulator in steady-state granulopoiesis, or may function as a regulator for emergency granulopoiesis in response to specific challenges requiring increased neutrophil production, such as infection (7).

G-CSF and its isolation, characterisation, and recombinant production have been extensively reviewed, for example papers cited in References 5 and 7.

Until recently, genetic studies depended upon the discovery of random mutations (either spontaneous or induced) or of pre-existing genetic polymorphisms. However, following the rapid development of recombinant DNA technology and of identification of specific genes, particularly in mice, by analogy to genes from other species or from the biochemistry of the protein products which they encode, methods for specifically-targeted deletion or modification of genes have been developed. Provided that a cloned, genomic fragment of the chosen genetic locus is available, it is possible to generate null alleles by disruption of the gene, to modify functional

properties of the gene such as transcriptional pattern, mRNA or protein maturation pattern, or to modify the ability of the protein to interact with other gene products. This is achieved by using conventional recombinant DNA methods to introduce the desired mutation into a cloned DNA sequence of the chosen gene; the mutation is then transferred by means of homologous recombination into the genome of a pluripotent embryonic stem cell (ES cell). The ES cells thus produced are transferred by microinjection into mouse blastocysts in order to generate germ-line chimeras. Animals homozygous for the desired mutation are then generated by interbreeding of heterozygous siblings.

These techniques are now widely used, and have been employed to generate lineages of mice in which a variety of genes are disrupted; these mice are often referred to as "knock-out" mice. Many of these mutations are lethal, causing death early in embryonic life or in the perinatal period. The technique has been particularly successful in producing "knock-out" mice in which genes for molecules of immunological importance or for growth factors are deleted. The techniques are well established, and a variety of marker genes and genes employed to assist in selection of cells which have undergone homologous recombination rather than random integration of DNA are available. A number of reviews have been published [7 to 11], and techniques for generation of transgenic mice in general are reviewed in International Patent Application No. WO 91/13150 by Ludwig Institute for Cancer Research.

While such gene targeting is useful in the production of mouse models for genetically-determined human diseases, which models can be used for testing potential therapies, and while the techniques are well established in principle, it is not possible to predict in advance whether an animal line bearing a given targeted gene disruption can be generated, or if so how readily practicable generation of such a model will be and the best experimental approach

to utilise. In particular, the frequency of transformation of the ES cells varies widely, from as little as 1 in 40,000 to as much as 1 in 150. While some of the factors involved in optimisation of transformation frequency are known, success is not easy to predict.

We have generated GM-CSF deficient and G-CSF deficient mice through targeted disruption of the GM-CSF and G-CSF genes respectively in embryonic stem cells. We have surprisingly found that while GM-CSF deficient mice have no major perturbation of haematopoiesis, they all have abnormal lungs and are prone to lung infections , implicating GM-CSF as being essential for normal pulmonary physiology and resistance to local infection. GM-CSF deficient mice are useful as a model system for the syndrome of alveolar proteinosis, and are particularly useful as a model system for the study of opportunistic infections and infections which are intractable to currently available therapies.

G-CSF deficient mice are neutropaenic, but have normal levels of monocytes in the periphery. They are prone to sub-clinical infections, and will be useful in testing the efficacy of anti-microbial agents, especially in settings of increased vulnerability. They will also be useful in assessing the virulence of microorganisms.

SUMMARY OF THE INVENTION

According to one aspect of the invention, there is provided a non-human animal carrying a disruption of a gene encoding a colony stimulating factor.

Preferably the animal is a rodent, for example a mouse, rat, rabbit or hamster, and more preferably is a mouse.

In one preferred embodiment, this aspect of the invention provides a non-human animal carrying a disruption of the gene encoding granulocyte-macrophage colony-stimulating-factor (GM-CSF), such that expression of GM-CSF

is disrupted. Preferably the spleen cells from the animal are incapable of producing detectable levels of bioactive GM-CSF.

Also preferably the gene encoding GM-CSF is completely inactivated. Most preferably the animal carries a mutation comprising deletion of exons 1 and 2 and intron 1 between *Sca*I and *Sma*I restriction sites of the gene encoding GM-CSF.

In an alternative embodiment, this aspect of the invention provides a non-human animal carrying a disruption of the gene encoding granulocyte colony-stimulating factor (G-CSF), such that expression of the colony-stimulating factor is disrupted. Preferably the lung cells from the animal are incapable of producing detectable levels of bioactive G-CSF. Also preferably the gene encoding G-CSF is completely inactivated. Most preferably the animal carries a mutation comprising a deletion from the *Nco*I restriction site in exon 1 to the *Bam*HI restriction site in exon 3 of the gene encoding G-CSF.

Optionally the animals may also carry one or more additional mutations which result in disruption of a specific gene. For example, a GM-CSF deficient animal may also carry a mutation resulting in disruption of the gene for G-CSF, or vice versa. Again the animal is preferably a mouse, and the additional gene disruption may alternatively be targeted to a gene encoding a growth factor selected from the group consisting of colony-stimulating-factor-1 (CSF-1); leukaemia inhibitory factor (LIF), and transforming growth factor-β1 (TGF-β1), or a cytokine, preferably an interleukin, including, but not limited to, an interleukin selected from the group consisting of interleukin-2, interleukin-3 and interleukin-6. Such double or multiple knock-out animals can be generated by crossing of animals in which the gene encoding GM-CSF or G-CSF is disrupted with animals in which the other desired gene(s) is disrupted. For example, the *osteopetrotic*

mouse, a line which originated by spontaneous mutation, is known to be deficient in CSF-1, and has decreased levels of monocytes/macrophages and of osteoclasts. Preferably the animal is a mouse deficient in both GM-CSF and G-CSF, in GM-CSF and CSF-1, or in both G-CSF and CSF-1.

Cell lines, such as bone marrow stromal cell lines, derived from the novel animals described herein are also within the scope of this invention.

According to a second aspect, the novel animals of the invention, especially mice, provide a convenient model system for the study of disease. In one embodiment, this aspect of the invention provides a model system for the syndrome of alveolar proteinosis, both of the congenital form and the adult form, in which potential methods for treatment of this condition can be tested.

In an alternative embodiment, this aspect of the invention provides a model system for the study of infectious diseases, especially bacterial infections of the lungs. In particular, this model system is useful for the study of opportunistic infections of the type to which immunocompromised or immunodeficient individuals are prone, and also of infections which are intractable to currently-available therapies, including antibiotic treatment. Immunocompromised individuals include persons suffering from cancer, especially those undergoing chemotherapy and/or radiotherapy, persons suffering from leukaemias, transplant recipients undergoing immunosuppressive therapy, and persons with autoimmune diseases undergoing immunosuppressive therapy. A particularly important group of individuals prone to opportunistic infections are those suffering from acquired immunodeficiency syndrome (AIDS).

The animals of the invention are expected to be useful for the study not only of bacterial and viral infections, but also infections with fungi and mycoplasma. Organisms suitable for study in the model system of the invention include, but are not limited to, *Listeria*, including *Listeria monocytogenes*, *Mycobacteria*, including

Mycobacterium tuberculosis, *Mycobacterium intracellulare*, *Mycobacterium avium*, *Mycobacterium bovis* and atypical *Mycobacteria*; *Pseudomonas* species, including *Pseudomonas aeruginosa*, Enterobacteriaceae, including *Escherichia coli*, *Salmonella* species, *Klebsiella* species, including *Klebsiella pneumoniae* and *Klebsiella oxytoca*; *Pneumocystis carinii*; *Histoplasma*, including *Histoplasma capsulatum*; *Cryptococcus*, including *Cryptococcus neoformans*; *Pasteurella*, including *Pasteurella pestis* and *Pasteurella pseudotuberculosis* (now known as *Yersinia pestis* and *Yersinia pseudotuberculosis* respectively), and *Pasteurella multocida* (also known as *Pasteurella septica*); and *Mycoplasma*, including *Mycoplasma pneumoniae*.

According to a third aspect of the invention, there is provided a method of treatment of pulmonary infection comprising the step of administering to a subject in need of such treatment an effective amount of GM-CSF or of G-CSF. Preferably the GM-CSF or G-CSF can be administered locally to the lung. This may be achieved by means of aerosols or nebulisers, for example ultrasonically nebulised formulations. Suitable carriers and excipients are known in the art. In intensive care situations, the CSF may be administered via an endotracheal tube. Optionally the GM-CSF or G-CSF may be administered as adjunctive therapy, for example concurrently with other known treatments for the particular condition. This aspect of the invention also provides a method of treatment of alveolar proteinosis or of pulmonary infection, comprising the step of administering to a subject in need of such treatment an effective amount of GM-CSF.

Where the GM-CSF or G-CSF is administered systemically, for example intravenously or subcutaneously, it is contemplated that dose ranges comparable to those known for use to stimulate granulopoiesis in patients undergoing cancer chemotherapy are suitable. Where GM-CSF or G-CSF is to be applied regionally to the lung, it is contemplated that the dose range will be from 0.1 μg/kg

body weight to 20 μg/kg body weight. The person skilled in the art will readily be able to determine suitable dose ranges by normal trial and error experimentation. In particular, the acceptable degree of toxicity at high doses will depend on the condition to be treated. Compositions for administration of GM-CSF or of G-CSF to the lung are also within the scope of the invention.

As will be discussed in detail below, we have surprisingly found that mice in which GM-CSF expression is disrupted are haematologically essentially normal, indicating that GM-CSF is not by itself required for development and maturation of granulocytes and monocytes/macrophages, at least in a steady-state situation. However, the increased susceptibility of the GM-CSF deficient mice to infection indicates that GM-CSF is required for "emergency" granulocyte and macrophage development. Similarly, mice in which G-CSF expression is disrupted are neutropaenic, but have essentially normal levels of monocytes/macrophages. Thus at least one other factor must be involved in steady state haemopoiesis, and the novel animals of the invention can be used to identify and isolate such factors.

Thus in a further aspect the invention provides a factor which is involved in regulation of steady-state haematopoiesis, and which is present in animals in which expression of GM-CSF or G-CSF is disrupted. Preferably the factor is also present in animals in which expression of both GM-CSF and G-CSF is disrupted, or in which expression of both GM-CSF and CSF-1 is disrupted.

As is well known in the art, such haematopoietic growth factors in mice have a high degree of homology with the corresponding factors in humans, and this homology is sufficient to enable a gene encoding a murine factor to be used as a probe for the isolation of the corresponding human factor. Even if the degree of homology is relatively low, iterative screening at low stringency can be used. Therefore this aspect of the invention also provides a gene

encoding a factor involved in regulation of steady-state haematopoiesis, and which is present in animals in which expression of GM-CSF or G-CSF is disrupted, which can be used for isolation of the corresponding gene. Preferably the gene from the animal has at least 30% homology to the human gene, more preferably at least 50% homology, and even more preferably at least 90% homology.

For treatment of individuals with alveolar proteinosis, including congenital alveolar proteinosis, gene therapy may be the most appropriate course. Methods for such gene therapy are known in the art, given that the identity of the defective gene is known and that the appropriate DNA has been isolated. In a particularly preferred form, it is contemplated that intravenous administration of liposomal formulations of cDNA encoding GM-CSF will be used, as described for example by N. Zhu, D. Liggitt and R. Debs *et al*, Science, 1993 261 209: "Systemic Gene Expression After Intravenous DNA Delivery Into Adult Mice", the contents of which are herein incorporated by reference.

Since it is evident from our results that genetically-determined deficiency of a colony-stimulating factor leads to increased susceptibility to infections, especially Gram-negative pneumonias, the invention also provides a method of diagnosis of a colony-stimulating factor deficiency, comprising the step of testing a tissue or cell sample from a subject suspected of suffering from such a deficiency for the absence of the gene encoding said factor. Preferably the colony-stimulating factor is GM-CSF, but G-CSF and M-CSF are also included within the scope of the invention. The test may suitably be carried out using peripheral blood lymphocytes, but may also use tissue obtained by biopsy, for example from lung. A test may be carried out using methods which are known *per se*, such as polymerase chain reaction, or reaction with a probe labelled with a detectable marker, for example using *in situ* hybridisation.

In a sixth aspect, the invention provides targeting constructs for disruption of the genes encoding GM-CSF and G-CSF respectively, as described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the scheme used for generation of GM-CSF deficient mice, and the strategy for disruption of GM-CSF gene, showing the intact GM-CSF gene with 4 exons (1,2,3&4), location of restriction enzyme sites (E=*Eco*RI, B=*Bgl*II, SC=*Sca*I, SM=*Sma*I, X=*Xmn*I, P=*Pst*I, H=*Hind*III, SL=*Sal*I), probe A (external to construct, diagnostic for targeted disruption) and probe B (corresponding to deleted sequences) and sites of PCR primer hybridisation (a=Ed120, b=Ed121, c=NEB#1224, see text). In the disrupted allele, exons 1 and 2 and intron 1 are deleted, replaced by the *E. coli* lac-z gene, human β-globin 3' untranslated and polyA addition sequences (βG) and PGK-*neo* (see text).

Figure 2 illustrates how deletion of GM-CSF exons 1-2 results in lack of GM-CSF immunoreactivity and bioactivity in spleen conditioned media. Southern blot of *Bgl*II digested tail DNA from PCR-genotyped GM+/+ and -/- mice probed first with probe A (see Figure 1A), confirming that deleted exons are present only in wild-type mouse, then reprobed without washing with probe B, confirming that the -/- mouse is homozygous for the disrupted GM-CSF allele and lacks the deleted sequences. Levels of immunoreactive (panel B) and bioactive (panel C) GM-CSF (solid columns) and bioactive IL-3 (open columns) in media conditioned by concanavalin-A and IL-2 stimulated splenocytes are shown for individual GM+/+ and GM-/- mice. In panel C, shaded column shows GM-CSF bioactivity after neutralisation with an anti-GM-CSF antibody.

Figure 3 shows typical appearances of lung histopathology in GM-CSF deficient mice:

A. Normal C57Bl/6 lung, central region (13 week, H&E X40)

B&C. GM-/- lung with moderate (B) and extensive (C)

lymphoid hyperplasia around central and peripheral vessels (11 week, H&E X40)

D. Detail of alveoli in GM-/- lung with large foamy macrophages, neutrophils and eosinophilic alveolar debris (7 week, H&E X400)

E-H. Immunoperoxidase staining of perivascular mononuclear cells in GM-/- mouse lung (16 week, X200) with primary antibodies [specificity]: E, PBS [negative control]; F, RA3-6B2 [B220]; G, GK1.5 [CD4]; H, 53.6-7 [CD8] (same nodule in each panel)

I-K. Focus of infection with fungal element in GM-/- lung (16 week, X400): I, positive control for Grocott stain; J, 5-10 μm Grocott-positive fungal particles; K, PAS-positive fungal particles in same location of contiguous section

L-N. Bacterial infections in GM-/- lungs: L, Gram stain control with Gram-positive and Gram-negative bacilli; M, Gram-positive coccobacilli in pneumonic consolidated area (7 week, X400); N, purulent acute *Pasteurella pneumotropica* lobar pneumonia in mouse dying at 4 weeks of age (H&E, X400)

O-Q. Features of 24 week GM-/- lung: granular refractile PAS-positive homogenous eosinophilc material in contiguous alveoli (O, H&E X400; P, PAS X400); Q, emphysematous area (arrowed) with persistent peribronchovascular lymphoid hyperplasia.

Figure 4 shows the ultrastructural appearance of lungs from GM-CSF deficient mice. Electron microscopic sections of 24 week old GM-/- lung (same mouse as Figures 3O-Q). Solid bars indicate 1μm.

A. Type II surfactant-producing alveolar cell with characteristic intracytoplasmic lamellar bodies. The adjacent alveolus contains Type-C lamellar body.

B. Numerous intra-alveolar Type-C lamellar bodies with characteristic "onion" appearance.

C. Intra-alveolar macrophage with phagosomes containing "onion" structures resembling Type-C lamellar bodies.

Figure 5a shows the screening strategy used to distinguish between wild-type and disrupted G-CSF alleles by genomic Southern analysis of *Xba*I digests of DNA. Hatched regions represent genomic sequences flanking the G-CSF gene. Black regions represent the exons of the G-CSF gene, and βG represents 3' untranslated sequences and polyA addition motif from the human β-globin gene.

Figure 5b illustrates the targeting construct, designated pKOGCSF3b used to generate the ES cell line 5.4 from which the G-CSF deficient animals were derived.

Figure 6 shows Southern blot analysis of the ES cell line 5.4, from which founding G-CSF+/- chimeras were generated. The probe used in the left hand panel hybridizes with a region outside the targeting vector, and is diagnostic for homologous integration of the targeting vector at one G-CSF allele. The probe for β-galactosidase, used in the right-hand panel, hybridises exclusively with a region within the disrupted allele.

Figure 7 shows the PCR-based screening strategy for the disrupted G-CSF allele. The primers are described in the text.

Figure 8 shows PCR analysis of DNA from tail tissue of G-CSF +/- chimeras and their progeny.

Figure 9 shows pups born from matings of heterozygous G-CSF+/- parents.

Figure 10 shows PCR analysis of tail tisuse DNA from G-CSF -/+ chimeras and their progeny, demonstrating germline transmission.

Figure 11 shows results of assays on conditioned media from organs of G-/-, GM-/- and wild-type mice.

Figure 12 shows that G-/- mice are unable to produce G-CSF. Results of conditioned media from organs of

G+/+ (wild-type) mice are compared with those of G-/- and GM-/- mice.

DETAILED DESCRIPTION OF THE INVENTION

The invention will now be described by way of reference only to the following non-limiting examples, and to the figures.

Abbreviations

The abbreviations used herein are as follows:

CSF	colony-stimulating factor
G-CSF	granulocyte colony-stimulating factor
GM-CSF	granulocyte-macrophage colony-stimulating factor
M-CSF	macrophage colony-stimulating factor
ES cells	embryonic stem cells
IL	interleukin
PCR	polymerase chain reaction
Taq polymerase	*Thermus aquaticus* DNA polymerase
ELISA	enzyme-linked immunoadsorbent assay
H&E	haematoxylin and eosin stain
PAS	Periodic acid-Schiff stain
SCM	spleen-cell conditioned medium/media
SCF	stem cell factor
G418	Geneticin (Gibco, Grand Island, NY)
bp	base pairs

Cell lines, vectors and antibodies

Cell lines used herein are as follows:

COS cells	cell line of monkey origin able to support replication of vectors comprising the SV40 origin of

	replication.
E14	ES cells derived from mice of strain 129/OLA (see Reference 17)
FDC-P1	cell line which responds to GM-CSF and IL-3
32-D	cell line which responds specificaly to IL-3

All these lines are well known in the art.

Vectors used herein are as follows:
pIC20H

Antibodies used herein are as follows:

MP1-22E9	specific for GM-CSF
MP1-31G6	second antibody used in ELISA
RA3-6B2	specific for B cells
187.1	specific for B cells
GK1.5	specific for T helper cells
59.6-7	specific for T suppressor cells
53-7.8	specific for pan T cell marker

MATERIALS AND METHODS

Statistics

Data are given as mean ± standard deviations. To test for statistically significant differences, the unpaired Student's t-test and Chi-squared test were used.

Generation of targeted ES cell colonies and chimeric mice

In their general aspects, the methods employed are known *per se*. Methods for isolation and injection of mouse blastocysts, and implantation into host females for generation of transgenic mice, are extensively reviewed in the volume "Manipulating the Mouse Embryo - A Laboratory Manual" by B. Hogan, F. Costantini and E. Lacy (Cold Spring Harbor Laboratory, 1986).

Example 1 **GM-CSF targeting vector and isolation of targeted ES cell clones.**

The GM-CSF targeting vector, illustrated in Figure 1 contained from 5' to 3' in plasmid pIC20H, approximately 900 bp of the GM-CSF promoter (*Rsa*I-*Sca*I) (13,14), the *E. coli* lac-z gene modified at the 5' end, an approximately 700 bp fragment of the 3' untranslated region and polyA addition motif from the human β-globin gene (*Eco*RI-*Acc*I), the PGK-*neo* selectable marker (15), and approximately 10 kb of GM-CSF genomic sequence from a previously isolated lambda clone (14). The targeting vector was constructed to delete GM-CSF exons 1 and 2 and intron 1 between *Sca*I and *Sma*I sites because in previous studies, cDNAs with this region deleted produced no GM-CSF activity when over expressed in COS cells (16). The vector was linearized at a *Sal*I site in the pIC20H polylinker prior to electroporation into E14 embryonal stem (ES) cells derived from mice of the strain 129/OLA (17), which were maintained and subsequently selected in medium containing the antibiotic G418, as previously described (18). After 8 days of G418 (Geneticin; Gibco) selection, individual G418-resistant colonies were cloned and replated.

The polymerase chain reaction (PCR) was used to screen DNA from each ES cell colony for integration of a targeting vector by homologous recombination. The PCR primers were:

"a" 5'-CCAGCCTCAGAGACCCAGGTATCC-3', corresponding to sequence 5' of the *Rsa*I site in the GM-CSF promoter;

"b" 5'-GTTAGAGACGACTTCTACCTCTTC-3', corresponding to sequence in exon 2 of the GM-CSF gene in the region deleted by targeted construct integration; and

"c" a M13 (-47) 24mer sequencing primer (New England Biolabs, Beverly, MA, #1224), corresponding to the 5' region of the *E. coli* lac-z gene.

These are illustrated in Figure 1. In PCR reaction mixtures, primers "a" and "b" generate a 1.2 kb product from wild-type DNA, and "a" and "c" a 1.0 kb product from DNA containing a correctly integrated targeted construct. PCR reaction mixtures (20μl) contained approximately 250ng DNA, 67mM Tris-HCl (pH 8.8), 16.6 mM $(NH_4)_2SO_4$, 0.45% Triton X-100, 200μg/ml gelatin, 1.5 mM $MgCl_2$, 250 μM of each deoxynucleoside triphosphate, 12.5ng of each primer and 1.5 U of Taq polymerase.

Following initial denaturation (95°C [150s]), 40 amplification cycles were performed (95°C [50s], 60°C [50s], 72°C [60s]). The structure of targeted alleles identified by PCR was confirmed by Southern blot analysis of DNA from tail tissue of mice homozygous for the disrupted GM-CSF allele, which were generated as described in Example 9 below. *Bgl*II-digested DNA was probed with a radiolabelled DNA fragment corresponding to GM-CSF genomic sequences which were deleted from the targeting construct (probe A, Figure 1, *Pst*I-*Pst*I fragment) to verify that these sequences were absent from the presumptive GM-CSF homozygous null mice, and reprobed with a probe B (Figure 1) corresponding to sequences lying outside the targeting construct (the *Bgl*II-*Xmn*I genomic fragment 5' of the *Rsa*I site in the GM-CSF promoter).

<u>Example 2</u> **Generation of Chimeric Mice.**

Two clonally independent ES cell lines (designated GM4 and GM6) with a targeted GM-CSF gene disruption were selected for injection into C57BL/6 mouse blastocysts. The resulting male chimeras from both cell lines transmitted the disruptive mutation through the germline in matings with females from a C57BL/6 background, to generate heterozygous male and female progeny. These heterozygous progeny were interbred to generate mice homozygous for the disruptive mutation. GM-CSF genotypes of mice have been designated as follows:

wild-type, GM+/+;

heterozygous GM+/-;
homozygous GM-/-.

The GM-CSF status of mice was routinely assessed by PCR analysis of DNA from tail tissue of the mice. GM-/- mice were subsequently bred from GM-/- x GM-/- matings, and to generate similarly outbred 129/OLA x C57BL/6 GM+/+ control mice, GM+/+ mice were bred from first- and second-generation GM+/+ littermates. Mice were kept in a conventional animal house. Except where stated, observations come from the GM4 lineage.

Example 3 **Verification of GM-CSF gene disruption**

To confirm the structure of the targeted allele in GM-/- mice, Southern blotting analysis was performed on *Bgl*II-digested tail DNA from randomly-selected GM+/+ and GM-/- mice which had been genotyped by PCR, using a probe corresponding to the exon 1 and 2 sequences deleted in the targeted GM-CSF gene. As expected, GM+/+ but not GM-/- tail DNA contained a 2.3 kb species hybridizing with the exon 1-2 probe. GM-/- DNA contained the predicted 7.7 kb species when hybridized with a probe corresponding to sequences lying outside the targeting vector, as shown in Figure 2A.

Example 4 **Spleen cell-conditioned media and GM-CSF assays.**

Spleen cell-conditioned media were prepared as previously described (19), except that the splenocytes were stimulated with concanavalin-A (5μg/ml, Boehringer-Mannheim, Germany) and interleukin-2 (100U/ml, Amgen, Thousand Oaks, CA). To assay immunoreactive GM-CSF, an ELISA was used with the antibodies MP1-22E9 and biotinylated MP1-31G6 (Pharmingen, San Diego, CA) with an avidin-biotinylated horseradish peroxidase (Dako, Copenhagen, Denmark) detection system with a sensitivity of 50 pg/ml. To assay bioactive GM-CSF, the proliferative response of FDC-P1 cells, measured by ^{3}H-thymidine

incorporation, was used, adjusting for interleukin-3 (IL-3) bioactivity on 32-D cells, using methods previously described (20). The MP1-22E9 neutralising antibody was used to confirm the specificity of the putative GM-CSF bioactivity. Standards used were recombinant murine GM-CSF (Schering-Plough, Kenilworth, NJ) and IL-3 (10^7 U/mg, Boehringer-Mannheim, Germany).

To confirm that tissues from GM-/- mice lacked GM-CSF, spleen cell-conditioned media (SCM) made from GM+/+ and GM-/- splenocytes were assayed for immunoreactive and bioactive GM-CSF. Using a GM-CSF ELISA, no immunoreactive GM-CSF was detected in SCM from GM-/- splenocytes while SCM from by GM+/+ splenocytes contained more than 2700 pg/ml of GM-CSF. These results are summarized in Figure 2B. To assay for bioactive GM-CSF, the proliferative response of FDC-P1 cells, which respond to GM-CSF and IL-3, was assayed in parallel to that of 32-D cells, which respond only to IL-3.

There was no detectable bioactive GM-CSF in SCM from GM-/- splenocytes, whereas GM+/+ SCM contained more than 100pg/ml bioactive GM-CSF which was at least 95% neutralised by an anti-GM-CSF antibody. To confirm the potency of these SCM, all were shown to contain bioactive IL-3, although SCM from GM-/- mice contained less IL-3 than did SCM from control mice (GM-/-, 16$\pm$2 U/ml, n=4; GM+/+, 49$\pm$20 U/ml, n=4; $p<0.01$). The inability of GM-/- tissues to produce GM-CSF *in vitro* was confirmed by studies on muscle and kidney conditioned media, using microwell assays of FDC-P1 cells (3).

<u>Example 5</u> Viability and fertility

From initial matings of GM+/- mice, litters of 10$\pm$3 pups (n=8, GM6) resulted, with the genotypes GM+/+, +/- and -/- represented in approximately Mendelian ratios amongst live births surviving to weaning, indicating that there was no selective fetal or neonatal loss of GM-/- pups. Survival of mice from initial litters was normal

amongst GM-/- pups (>91%, n=35) compared to their GM+/+ littermates (>88%, n=17), with a median follow-up for GM-/- mice of 220 days [range 0-334] days and for GM+/+ mice of 209 days [range 0-313] days; median survival has not been attained. *Post-mortem* examination of the two dead GM-/- adults in this cohort indicated that one, which died at 153 days, had lymphoid leukaemia and the other, which died at 167 days, had hepatitis; both displayed the characteristic GM-CSF deficiency lung disease which is described in detail below. From initial matings of male and female GM-/- mice, litters of 9±1 pups (n=5) resulted, indicating that GM-CSF deficiency did not grossly impair fertility or fecundity.

Example 6 Haematological analysis of GM-CSF deficient mice

Haemoglobin, total leukocyte and platelet estimates were performed on 1:4 dilutions of eye-bleed samples on a Sysmex-K1000 automated counter, and manual 100-cell leukocyte differential counts were performed on May-Grunwald/Giemsa stained smears.

Progenitor cells were assayed in semisolid agar cultures of bone marrow, spleen or peritoneal cells by a standard method, as previously described (19,21). Colony formation was stimulated by purified recombinant bacterially-synthesized growth factors at the following final concentrations: human G-CSF [10 ng/ml], murine GM-CSF [10 ng/ml], murine IL-3 [10 ng/ml], murine M-CSF (CSF-1) [10 ng/ml] rat stem cell factor [100 ng/ml], murine IL-6 [500 ng/ml], and spleen cell-conditioned medium (10%). Colonies were typed on stained whole-plate preparations as previously described (21).

The haematological parameters of peripheral blood of 6-7 weeks GM-/- mice showed no significant difference from those of blood of GM+/+ littermates, as illustrated in Table 1.

Table 1

Haematological Analysis of GM-CSF - deficient mice

	GM+/+	GM-/-
Haemoglobin (g/l)	162±7	163±5
total leukocytes ($x10^9$ cells/l)	5.9±1.0	7.4±2.4
neutrophils ($x10^9$ cells/l)	1.1±0.3	1.2±0.6
lymphocytes ($x10^9$ cells/l)	4.7±1.1	6.0±2.0
monocytes ($x10^9$ cells/l)	0.12±0.10	0.13±0.13
eosinophils ($x10^9$ cells/l)	0.09±0.06	0.13±0.13
platelets ($x10^9$ cells/l)	838±105	822±109

10 mice in each group were tested

Results are expressed as mean ± S.D.

GM-/- mice tended to have greater variation in their granulocyte levels, as illustrated by the granulocyte levels of 5-7 week mice: GM-/- 1.7±1.5 x 10^9/l [n=33, range 0.2-6.6]; GM+/+ 1.3±0.7 x 10^9/l [n=15, range 0.29-3.1], with 24% of GM-/- mice having neutrophils greater than 2 x 10^9/l compared to 15% of GM+/+. Spleens of GM-/- mice showed increased variability in mass (e.g. spleen mass of 6 weeks mice, n=6/group: GM+/+ 106±9 [range 94-120] mg; GM-/-, 114±42 [range 64-191] mg).

Haematopoiesis was evaluated in five pairs of 6-8 week GM+/+ and -/- mice. Femoral cellularity was equivalent (GM+/+ 34.0±5.3 x 10^6, GM-/- 27.4±7.0 x 10^6 cells/femur, n=3) and the myeloid:erythroid ratios were equivalent (20±2 and 17±6 % erythroid cells respectively). The results of analysis of marrow progenitor cells responsive to various stimuli, summarized in Table 2, revealed no major difference in total progenitor cell frequency, and colony subtyping indicated no differences in frequencies of granulocyte, granulocyte-macrophage,

macrophage, eosinophil, megakaryocyte, erythroid and blast marrow progenitor cells. To assess the role of GM-CSF in the phenomenon of "spontaneous" *in vitro* colony formation, crowded unstimulated cultures of up to 2 x 10^5 GM-/- marrow cells/ml were established; colony formation was somewhat reduced, indicating that "spontaneous" colony formation is not solely dependent on *in vitro* GM-CSF production.

Table 2

Haematopoietic progenitor cells in GM-CSF deficient mice

Cells cultured (n)	Stimulus	Total colonies (mean ± S.D. GM+/+	GM-/-
25,000 bone marrow (5)	GM-CSF	59±12	51±19
	G-CSF	19±3	15±7
	M-CSF	64±16	59±25
	IL-3	68±12	68±19
	SCF	29±3	22±13
	IL-6	18±8	18±10
	SCF + GM-CSF	66±12	63±26
200,000 bone marrow (5)	Saline	79±62	39±44
100,000 spleen (3)	GM-CSF	3±4	8±2
	SCF	0	4±2*
	SCF + GM-CSF	3±1	20±2*
	SCM	4±3	18±2*

(n) indicates the number of individual mice studied
* indicates $p<0.001$.

Cultures were stimulated by: granulocyte-macrophage-CSF (GM-CSF), granulocyte-CSF (G-CSF), CSF-I (M-CSF), interleukin-3 (IL-3), stem cell factor (SCF), interleukin-6 (IL-6), spleen cell-conditioned medium (SCM), or a combination of these.

Genotype is indicated as wild-type (GM+/+) or GM-CSF deficient (GM-/-).

The spleen masses for these groups of mice were: GM+/+, 64±8 mg; GM-/-, 114±22 mg ($p<0.001$).

Analysis of splenic haematopoiesis showed a 3 to 6 fold increase in frequency of progenitor cells in GM-/- mice, and as the GM-/- spleens were significantly larger, this represented an absolute increase in number of splenic progenitor cells (Table 2). Peritoneal washings recovered $6.0 \pm 1.4 \times 10^6$ and $5.1 \pm 1.4 \times 10^6$ cells from GM+/+ and GM-/- mice respectively; these washings comprised 65% macrophage in GM+/+ and 63% macrophages in GM-/- mice.

<u>Example 7</u> **Histological examination of GM-CSF deficient mice and characterisation of pulmonary disease**

Formalin-fixed paraffin-embedded sections of mouse organs were stained using standard techniques with haematoxylin and eosin (H&E), and selected sections were stained with the Grocott methenamine silver and periodic-acid-Schiff (PAS) stains, and Ziehl-Neelsen and Wade-Fite acid fast stains. Immunoperoxidase staining of lung tissue was performed using a standard method (22) on 4 μm frozen sections, with the following antibodies: RA3-6B2 [specific for B220] (23), 187.1 [specific for κ light chain] (24), GK1.5 [specific for CD4] (25), 53.6-7 [specific for CD8] (26), and 53-7.8 [specific for CD5] (Pharmingen, San Diego, CA). For electron microscopy, random strips of fresh lung tissue were immersion-fixed in 2.5% glutaraldehyde in 1% sodium cacodylate buffer (pH 7.4), postfixed in 2% aqueous osmium tetroxide, and embedded in Araldite-Epon resin using standard techniques. Thin sections were stained with alkaline lead citrate and uranyl acetate, and viewed in a Jeol 1200EX electron microscope.

Histological sections from GM-/- mice aged 1 day to 24 weeks and age-matched GM+/+ controls of age were reviewed. No major differences between GM+/+ and GM-/- were evident in haematopoietic tissues; marrow cellularity was normal and some GM-/- spleens were enlarged with hyperplasia of both red and white pulp. Whilst at birth the lungs of GM-/- and GM+/+ animals were

indistinguishable, within the first 3 weeks of post-natal development striking abnormalities were evident in the lungs of GM-/- mice. Individual lungs from GM-/- mice consistently showed focal aggregates of lymphoid cells in peribronchial and perivascular areas, but little infiltration of alveolar septa; typical results are shown in Figure 3A-C. Immunoperoxidase staining of 12-16 week GM-/- lungs showed these cells to be predominantly B-lymphocytes on the basis of positive staining for B220 and κ-light chain. About 20% of lymphocytes were T-cells, predominantly of the $CD4^+$ type (Figure 3E-H). The lymphoid foci were particularly marked around large hilar vessels, occasionally assuming a follicular organization, but the cells exhibited little mitotic activity. The foci characteristically extended peripherally around small bronchioles and arterioles.

Focal consolidation was prevalent, and consisted of an intensely eosinophilic alveolar exudate containing numerous mature and fragmented neutrophils and macrophages (e.g Figure 3D,M,N). Associated bronchioles contained purulent exudate. In almost every GM-/- mouse of up to 6 weeks of age, acute inflammatory areas were found in the distal tips of the pulmonary lobes. Lobar consolidation was frequently observed (Figure 3M&N). In lungs of older mice (6-12 weeks), the lymphoid hyperplasia predominated; alveoli contained large foamy macrophages and granular debris, and foci of acute inflammation, comprising aggregates of up to 50 neutrophils and mononuclear cells, were sometimes seen (Figure 3D). Granular, eosinophilic, PAS-positive material within alveoli was present in all lungs examined (e.g. Figure 3D&K), apparently accumulating and becoming confluent in some alveoli by as early as 3 weeks. This material was present until at least 6 months of age, and in some areas of 12-24 week GM-/- lungs contiguous alveoli containing this material resembled the appearances of alveolar proteinosis (Figure 3O&P).

Ultrastructurally, surfactant-producing type-II alveolar cells were readily identified by their characteristic cytoplasmic lamellar bodies (Figure 4A); the alveolar debris included numerous type-C lamellar bodies (Figure 4B), and these onion-like bodies were seen within phagosomes of intra-alveolar macrophages (Figure 4C).

Peripherally, the alveolar spaces of lungs from older mice (24 weeks) were large, suggesting an emphysematous process (Figure 3Q). One 4 week-old GM-/- mouse died with florid lobar pneumonia (Figure 3N) from which *Pasteurella pneumotropica* was isolated, corresponding to Gram-negative organisms evident in lung sections. A survey of GM-/- lungs with Grocott and PAS stains identified foci of 5-10 μm diameter fungal elements in 3/15 GM-/- lungs but 0/7 GM+/+ lungs (e.g. Figure 3I-K). Surveys of sections with Gram stain identified large numbers of Gram-positive coccobacilli in one pneumonic area (Figure 3L&M). No mycobacterial infections were evident with Ziehl-Neelsen and Wade-Fite stains. One 6 week-old GM-/- mouse had developed a chronic pulmonary abscess with an organized wall lined by foamy macrophages and containing neutrophil-rich pus.

<u>Example 8</u> **GM-CSF Deficient Mice Can Be Used to Test Host Response to Infection**

Ten wild-type and ten GM-CSF deficient mice were challenged by intravenous injection with *Listeria monocytogenes*, at a dose of 5 x 10^3 colony forming units per mouse. Haematological parameters and bacterial counts in the liver and spleen of mice sacrificed at day 1 and day 5 after infection are summarised in Table 3.

Table 3
Haematological Response to Listeria Infection

Haematological Parameter	Before injection		Day 1		Day 5	
	GM+/+ (10)	GM-/- (10)	GM+/+ (10)	GM-/- (10)	GM+/+ (10)	GM-/- (10)
Haemoglobin g/L	162±7	163±5	148±9	152±13	134±6	127±6
Platelets x10^9/L	838±105	822±109	750±96	710±99	776±375	352±340
Total White Cells x10^9L	5.9±1.0	7.4±2.4	10.9±3.5	8.3±2.0	11.8±1.6	5.2±2.2
neutrophil	1.1±0.3	1.2±0.6	4.6±1.4	1.5±0.9	3.4±1.1	1.14±0.55
lymphocytes	4.7±1.1	6.0±2.0	6.3±2.3	6.0±1.4	7.3±0.9	3.5±2.1
monocytes	0.12±0.10	0.13±0.13	0.18±0.25	0.12±0.13	1.11±0.18	0.55±0.25
Eosinophils	0.09±0.06	0.13±0.13	0.11±0.13	0.12±0.15	0.03±0.06	0

Listeria organism counts (log mean ± SD)

		Day 1		Day 5	
		GM+/+	GM-/-	GM+/+	GM-/-
		(5)	(5)	(5)	(5)
Spleen	per organ	5.00±0.48	5.11±0.35	2.65±0.10	5.98±1.44
	per gm	6.05±0.45	6.11±0.40	3.26±0.18	6.76±1.66
Liver	per organ	4.17±0.51	5.02±0.40	2.04±0.41	6.15±1.92
	per gm	4.24±0.51	5.03±0.46	1.96±0.39	6.16±2.00

Results for the haematological parameters are presented as mean ± standard deviation, while those for organism counts are presented as log mean ± standard deviation. The figures in brackets represent the number of mice in each group.

On Day 5, 2 GM-/- mice were dead, with organism counts of greater than 10^8 per organ for spleen or liver. The surviving GM-/- animals had 5.66 ± 0.95 organisms g/spleen and 4.81 ± 1.07 organisms/liver, which was still much higher than the counts for GM+/+.

Haematologically, the GM-/- mice showed a perturbed response, with an impairment of neutrophil mobilisation on Day 1 and neutrophil production on Day 5, despite the presence of estabilshed *Listeria* infection. The GM-/- mice were unable to control the *Listeria* infection as well as could their littermate controls, showing a higher bacterial load in the liver on Day 1, and a higher bacterial load in both liver and spleen on Day 5. Thus the GM-CSF deficient mice do constitute a useful model for testing the host response to infection.

Example 9 **GM-CSF Deficient Mice Are A Suitable Model For Vulnerability To Infection**

Blood counts on age-matched 3 week-old GM-/- mice reared under conventional conditions were compared with those from mice reared in a microisolation ("specific pathogen-free"; SPF) environment. The results are summarised in Table 4.

Table 4

Comparison between GM-/- Mice Reared in Conventional and SPF Environments

Parameters	SPF	Conventional	p Value (t test)
Whole body mass (g)	13.4±0.8	11.9±2.1	NS
liver mass (mg)	576±63	576±95	NS
spleen mass (mg)	72±26	95±8	<0.1 (NS)
spleen mass % body wt	0.53±0.16	0.87±0.16	<0.02
Haemoglobin g/L	135±6	148±9	<0.05
Total White Cells x10^9/L	7.7±2.2	8.9±3.4	NS
Neutrophil	0.99±0.70	2.98±2.02	(NS) <0.1
Lymphocytes	5.55±2.12	5.23±1.01	NS
Monocytes	0.13±0.06	0.69±0.50	<0.05
Eosinophils	0±0	0±0	NS
Platelets x10^9/L	763±91	804±16	NS

NS = not significant

The mice from the microisolation environment had lower neutrophil and monocyte levels, strongly suggesting that the mice from the conventional environment had been exposed to a range of infectious agents during their first three weeks of life. Despite the lack of statistical significance as assessed by Student T-Tests, there is a marked indication that the mice from the microisolation environment had lower neutrophil levels and a profile suggestive of lesser degree of infection. Thus there is strong evidence for an environmental influence on these parameters.

Discussion

Since the actions of various haematopoietic regulators on target cells appear to overlap, it is possible that individual regulators might be wholly or partially redundant (27). This proposition is most directly assessed by analysis of mice deficient in individual regulators or combinations of regulators. Our analysis of GM-CSF deficient mice up to 3 months of age found that there was no perturbation of major haematopoietic populations in marrow or blood. There are two obvious ways in which to interpret this observation: either GM-CSF may not be a pivotally important regulator of normal haematopoiesis, or alternatively, GM-CSF may contribute to the maintenance of steady-state haematopoiesis, but in its absence other haematopoietic regulators are able to replace the role normally carried out by this regulator. This latter possibility would be further supported if elevated levels of compensatory regulators were demonstrated in the GM-CSF deficient mice. Splenic progenitor cell levels were increased in GM-CSF deficient mice, but this may in part reflect subclinical pulmonary infection in the animals; it will be important to assess to what extent the pulmonary disease process is impacting on "steady-state" haematopoiesis, particularly in the spleen.

The possibility that GM-CSF is a wholly redundant molecule can clearly be discounted in the light of our observation that all GM-CSF deficient mice show abnormal lung development. Our initial studies have not identified the nature of the intrinsic pulmonary defect. However, the same pathology is not seen in control animals, strongly suggesting that it is due to GM-CSF deficiency. A prominent pathological feature is granular eosinophilic intra-alveolar material, which sometimes becomes confluent. This may be a local product, such as pulmonary surfactant phospholipid and protein, either produced in excess or cleared too slowly due to a functional defect of macrophages. The presence of numerous type-C lamellar bodies within alveoli and macrophages is consistent with the accumulation of surfactant components (28). In some areas, the histological appearance of lungs from GM-CSF deficient mice bears a striking resemblance to some forms of alveolar proteinosis, a heterogeneous group of congenital and acquired lung disorders characterised by the intra-alveolar accumulation of surfactant protein, and often complicated by infection (28-31). The roles of GM-CSF or of cells functionally activated by GM-CSF in the production of surfactant by type II alveolar cells and on surfactant clearance have not been studied. Whilst clearance of mucinous secretions in larger airways may also be defective, retained secretions are not prominent in the lungs of younger GM-CSF-/- mice, which already have excessive alveolar material.

A prominent feature of the lung pathology of GM-CSF deficient mice is the presence of infection with a range of opportunistic organisms, including both bacterial and fungal agents. Interestingly, infection of immunocompromised CD4+ T-cell depleted mice with *Pneumocystis carinii* results in peribronchovascular lymphoid hyperplasia of similar appearances to that in GM-/- mice (32), suggesting that the lymphoid hyperplasia may be an aspect of the general pulmonary response to

infection. When infection occurs, the host response is usually adequate to prevent death, as only less than 0.5% of GM-/- mice have died from lung infection. However, the lymphoid hyperplasia, prevalent acute bacterial pneumonia, and subclinical multifocal fungal infection suggest that the host response to infection is defective. *Pneumocystis carinii* pneumonia is common in AIDS patients, frequently being the ultimate cause of death.

The presence of large numbers of neutrophils and macrophages within GM-/- lungs indicates that inflammatory cells can still be localized in the lung in the absence of GM-CSF to activate acute and chronic inflammatory processes within tissues, although the numerical and functional adequacy of the cells involved in this response has not been evaluated. Indeed, alveolar macrophages are particularly responsive to GM-CSF, and many cell types present in the lung (including epithelial cells) are capable of GM-CSF synthesis (33,34). It is possible, therefore, that a significant component of the intrinsic pulmonary defect is the absence of local GM-CSF-dependent activation of macrophages involved in either surfactant clearance or infection control. We are testing the influence of environmental factors on the pulmonary manifestations of GM-CSF deficiency by comparing GM-CSF deficient mice raised in conventional and gnotobiotic animal facilities.

The prominent pulmonary pathology accompanying absolute GM-CSF deficiency suggests that GM-CSF may be useful for treatment of lung disorders characterised by accumulation of alveolar material such as alveolar proteinosis, or by infection. Acquired forms of alveolar proteinosis may reflect a local relative deficiency of GM-CSF. Amongst congenital forms of alveolar proteinosis, there may be a human counterpart to murine GM-CSF deficiency for which GM-CSF replacement therapy would be appropriate.

There is considerable evidence that GM-CSF can be effective in treatment or prevention of bacterial and viral infections, and may also have some role in resistance to parasite infection (34-38).

<u>Example 10</u> **G-CSF Targeting Vector And Isolation of Targeted ES Cell Clones**

The G-CSF targeting vector, illustrated in Figure 5, was designed to effect a disruption of the G-CSF gene (Tsuchiya, M., Kaziro, Y. and Nagata, S.; Eur. J. Biochem., 1987 <u>165</u> 7-12) comprising a deletion of sequences between the *Nco*I site in exon 1, which encompasses the translational initiation codon ATG, and the *Bam*HI site in exon 3. Sequences comprising the *E. coli* lac-z gene with a modification at the 5' end to facilitate cloning into a *Nco*I site, and the neomycin phosphotransferase gene driven by the phosphoglycerate kinase (PGK) promoter, and with 3' untranslated sequences derived from the vector pMClNeo were inserted at the site of this deletion.

The construct, designated pKOGCSF3b, comprised a hybrid vector backbone derived from the vectors pBR322 and pSP73, from 5' to 3', approximately 1.1 kb of genomic sequence immediately 5' of the G-CSF protein coding sequence, including the G-CSF promoter (the Asp 700 - *Nco*I fragment), the *E. coli* lac-z and neomycin phosphotransferase sequences set out above, and approximately 7 kb of genomic sequence commencing at the *Bam*HI site in exon 3 and limited at the 3' end by a *Bam*HI site. This construct is illustrated in Figure 5b.

In this figure the approximate sizes in kb of genomic sequences are indicated by the bars. The *Nco*I site labelled with the asterisk encompasses the translation initiation site in exon 1 of the murine G-CSF gene. The *Bam*HI site labelled with a cross lies in exon 3 of this gene. βG represents a 700 bp fragment of the 3'

untranslated region and polyA addition motif from the human β-globin gene. The sequences in the neomycin phosphotransferase expression cassette are derived from PGK-Neo (sequences 5' of the *Nco*I site) and pMC1Neo (sequences 3' of the *Nco*I site). The vector backbone is a hybrid of sequences derived from pBR322 and pSp73. The sequences from these two vectors are joined in the targeting construct at the *Xmn*I site in the *Amp*^{R} gene to recreate a functional *Amp*^{R} gene. The origin of replication is derived from pBR322. The 3' end of the homologous genomic sequence in the construct is defined by a *Bam*HI site in genomic sequences, although this site was ablated during construction of the targeting vector.

Transformation and selection of ES cells were carried out generally as described in Example 1. Polymerase chain reaction was used to screen DNA from each ES cell colony for intergration of the targeting vector by homologous recombination. The PCR primers were as described below.

Chimeric mice were generated and maintained as described in Example 2. Except where stated, the following observations were obtained using the 5.4 lineage.

__Example 11__ Verification Of G-CSF Gene Disruption

To confirm the structure the targeted allele in G-/- mice, Southern blotting analysis was carried out on genomic DNA from ES cell lines or mouse tail tissues, which was prepared by standard techinques and digested with *Xba*I. Results are shown in Figure 6. This figure confirms the predicted structure of the disrupted G-CSF allele, using probes corresponding to sequences both external to and internal to the construct. The probe corresponding to G-CSF genomic sequences lying outside those of the targeting construct pKOGCSF3b was a 5'-*Sac*I-*Bgl*II fragment. This probe was diagnostic for homologous integration at one G-CSF allele.

DNA from the ES cell line 5.4, which as described

below was confirmed to contain a disrupted G-CSF gene, showed a 14 kb band corresponding to the wild-type allele, as well as a second larger band corresponding to the disrupted allele. The identity of this second band was confirmed by reprobing the Southern blot with a second probe corresponding to 3.2 kb of sequence from the *E. coli* lac-z gene, showing that only a band corresponding in size to that of the disrupted allele contained lac-z sequences. Thus incorporation of these sequences into genomic DNA at the disrupted G-CSF allele was confirmed, as well as the fact that only one copy of the targeting construct sequences was present in the genome of these ES cells.

<u>Example 12</u> **PCR Analysis**

A PCR probe specific for the disrupted G-CSF allele was used to demonstrate that matings of heterozygous parents resulted in progeny which were homozyous for the disrupted G-CSF allele. These results are shown in Figure 7. G105 template DNA was shown to have the genotype G+/- by Southern analysis. The symbols -, + and ++ indicate primer concentrations of 0.20 ng/20 µL reaction mixture, and 40 ng/20 µL reaction mixture. The primer sequences are:

GL#2	5'-TCC.ATG.ACA.TCT.GTG.GCC.ACC.AGA-3'
ES#132	5'-CTG.CAA.GGC.GAT.TAA.GTT.GGG.TAA-3'
GL#4	5'-CGG.CCT.CTC.GTC.CTG.ACC.ATA.GTC-3'

The PCR reaction mixtures were as described in Example 2, except that the concentration of $MgCl_2$ was 2.0 mM and 20 ng of each primer was used, and 35 amplification cycles were performed, annealing at 62° for 50 seconds.

When maximal sensitivity was required, PCRs to identify wild-type and disrupted alleles were performed separately. When DNA from tail tissue was screened, PCRs to diagnose wild-type and disrupted allelic status were carried out in a single reaction mixture, which contained 40 mg of primer GL#2 and 20 ng of the other two primers.

Primers GL#2 and GL#4 generate a 1.2 kb PCR product from template DNA containing a wild-type allele, while primers GL#2 and ES#132 generate a 1.1 kb PCR product from template DNA containing a disrupted GSF allele.

Results of PCR analysis on the first litter resulting from mating G-CSF +/- parents are shown in Figure 8. The PCR pattern demonstrates that progeny homozygous for the disrupted G-CSF allele were produced. We have thus demonstrated germline transmission of the disrupted allele. We have also shown that animals bearing the disrupted allele are viable and fertile.

__Example 13__ **Cells From G-CSF Deficient Mice Are Unable To Produce G-CSF**

An assay specific for mouse G-CSF was devised and validated. This assay utilised BAF3 cells modified to express receptors for G-CSF. These cells are designated BAF3-GR. Control (untransformed) BAF3 cells, which respond to Multi-CSF and also respond weakly to IL-2, were used as controls. Conditioned medium from thymus, lung, bladder, bone shaft and spleen were compared with a control preparation containing 1 mg/mL of purified G-CSF. These results are shown in Figure 11, and indicate that the BAF3-GR cells respond specifically to G-CSF, with a lower limit of sensitivity of 30 pg/mL. The results obtained with BAF3 cells show that neither Multi-CSF nor IL-2 is present in the conditioned medium from these organs. Conditioned media from thymus, lung, bladder, bone shaft and spleen were prepared from organs of G-CSF deficient and wild-type animals, and assayed in the BAF3-GR test system. All of the conditioned media from organs of wild-type mice contained titratable levels of bioactive G-CSF, whereas no bioactive G-CSF could be detected in conditioned media from organs of G-/- animals. The results are illustrated in Figure 12. In addition, assays of sera of G-/- and wild-type mice were unable to detect G-CSF. With serum the lower limit of sensitivity is 120pg/ml.

Example 14 **Haematological Analysis of G-CSF Deficient Mice**

Haemoglobin, total leukocyte and platelet estimates were performed on samples from G-CSF deficient mice, as described in Example 6. The results are summarised in Table 5, and show that the G-CSF deficient mice (-/-) have a durable base line neutropaenia compared to wild-type (+/+) or heterozygous (+/-) mice. However, in the G-CSF deficient mice other parameters were within the normal range.

Table 5
Haematological Parameters in G-CSF -/- Mice

Mice aged 3-4 weeks	Genotype (n)		
	G+/+ (31)	G+/- (30)	G-/- (18)
Haemoglobin g/L	142±12	140±13	140±15
Platelets $x10^9$/L	514±222	519±199	532±198
Total White Cells $x10^9$/L	6.79±1.66	6.25±2.09	5.48±2.16
Neutrophils $x10^9$/L	0.63±0.29	0.42±0.27	0.17±0.12
Lymphocytes x109/L	5.59±1.84	5.45±2.05	5.04±2.07
Monocytes $x10^9$/L	0.27±0.19	0.22±0.15	0.23±0.13
Eosinophils	0.14±0.12	0.13±0.13	0.10±0.10

Mice aged 12-16 weeks	G+/+ (5)	G+/- (0)	G-/- (7)
Haemoglobin g/L	169±9	ND	169±18
Platelets $x10^9$/L	798±134		802±125
Total White Cells $x10^9$/L	10.0±2.5		7.66±3.0
Neutrophils	0.57±0.19		0.20±0.12
Lymphocytes	9.10±2.37		7.24±2.94
Monocytes	0.28±0.19		0.17±0.05
Eosinophils	0.07±0.08		0.03±0.04

n = number of mice
Results are expressed as mean ± s.d.

Example 16 **G-CSF Mice Constitute A Model Of Chronic Neutropaenia**

The durability of neutropaenia in the G-CSF deficient mice is evident from the data presented in Example 13. A model of chronic neutropaenia is useful to test new therapies applicable to diseases characterised by chronic neutropaenia, and to test the role of neutrophils in experimental models infection, inflammation or malignancy, in a situation of relative neutrophil deficiency.

For example, a 154 day-old female G-/- mouse was killed because it appeared ill, and was found to have a severe endometritis accompanied by enlarged para-aortic lymph nodes, and enlargement of spleen, lungs, heart and liver. Histological examination confirmed pelvic inflammation and endometritis, as well as presence of large numbers of intensively eosinophilic, multinucleated macrophages in the alveoli of the lungs. This mouse had a severely abnormal blood profile (haemoglobin 108 g/L, total leukocyte count 29 x 10^9/L, and platelet count 418 x 10^9/L), and 70% of the circulating leukocytes were monocytes. Endometritis is an unusual infection in mice, and we have not seen it in any other immunocompromised animal in the last twelve months. The infection was probably introduced during mating, and the deficiency of neutrophils and G-CSF probably contributed significantly to the vulnerability of the animal to this infection.

The profound, excessive monocytosis seen in this animal during bacterial infection is very abnormal, and it appears that inability to make G-CSF had impaired its capacity to mount an appropriate granulocytic response to infection. This monocytosis may indicate an attempt to compensate for the lack of neutrophils by generating monocytes and macrophgaes via the remaining intact mechanisms of haemopoiesis.

Example 17 **G-CSF Deficient Mice Are Vulnerable To Bacterial Infection**

G-CSF deficient and wild-type mice were challenged with *Listeria monocytogenes*, administered by intravenous injection at a dose of 5 x 10^3 colony forming units per mouse, and haematological and bacterial load end points were examined at 1 hour, 1 day and 5 days after injection. The results are summarised in Tables 6 and 7.

Table 6

Haematological Parameters in Mice Injected with *Listeria*

Haematological Parameters	Baseline		1hr after *Listeria*	
	+/+	G-/-	+/+	G-/-
	(13)	(13)	(13)	(13)
Haemoglobin g/L	171±8	159±8	163±9	156±6
Platelets $x10^9$/L	843±122	759±118	787±149	773±126
Total White Cells $x10^9$/L	9.3±1.7	6.0±1.6	6.2±1.5	3.2±1.0
Neutrophils	1.05±0.39	0.18±0.10	2.85±1.39	0.30±0.17
Lymphocytes	7.56±2.12	5.62±1.41	3.18±0.38	2.76±0.99
Monocytes	0.15±0.09	0.14±0.13	0.07±0.08	0.09±0.08
Eosinophils	0.04±0.10	0.04±0.06	0.06±0.06	0.04±0.05

Table 6 (cont.)

Haematological parameter	1 Day After *Listeria* Injection		5 Days After *Listeria*	
	+/+	G-/-	+/+	G-/-
	(13)	(13)	(8)	(8)
Haemoglobin g/L	157±8	155±4	140±11	134±8
Platelets $x10^9$/L	729±134	772±143	960±251	663±464
Total White Cells $x10^9$/L	10.83±2.97	5.20±1.60	10.35±1.41	3.84±2.05
Neutrophils	4.28±2.31	0.77±0.43	2.70±1.26	0.33±0.27
Lymphocytes	5.72±1.80	4.16±1.67	6.88±2.10	3.07±2.22
Monocytes	0.65±0.32	0.23±0.12	0.66±0.30	0.45±0.39
Eosinophils	0.17±0.13	0.08±0.06	0.05±0.08	0±0
FACS Phenotype (m)	(8)	(8)	(7)	(7)
Mac^+ Grl^+ %	53	21	42	17
Mac^+ $F4/80^+$ %	ND	ND	12	19
Mac^+ Grl^{low} %	ND	ND	13	19
Mac^+ Gr^- %	7	9	ND	ND

Figures in brackets = number of mice

FACS = fluorescence activated cell sorter

Table 7

Response of G-/- Mice to *Listeria* Infection

	Bacteria Load			
	Spleen		Liver	
Wild-type (mouse no.)	Bacteria/g	Bacteria/spleen	Bacteria/g	Bacteria/liver
17	4.8×10^4	1.1×10^4	9.3×10^4	1.6×10^5
18	1.1×10^5	3.3×10^4	2.9×10^4	4.3×10^5
19	5.8×10^5	2.6×10^5	3.9×10^4	6.9×10^4
20	8.8×10^4	3.7×10^4	3.5×10^4	6.4×10^4
24	5.6×10^4	2.0×10^4	4.8×10^3	6.1×10^3
25	2.7×10^5	8.2×10^4	5.2×10^4	5.4×10^4
26	1.9×10^5	8.0×10^4	1.0×10^5	1.5×10^5
23	1.1×10^7	2.4×10^6	8.3×10^4	1.0×10^5
G-/- (mouse no.)				
30	3.4×10^5	1.1×10^5	1.4×10^4	2.3×10^4
33	1.7×10^5	5.4×10^4	8.1×10^4	1.6×10^5
37	2.0×10^9	3.0×10^8	2.6×10^8	2.8×10^8
38	4.1×10^8	5.4×10^7	1.3×10^7	1.4×10^7
39	8.1×10^6	2.4×10^6	3.5×10^5	4.8×10^5
36	$>2.4 \times 10^9$	$>3.0 \times 10^8$	$>5.7 \times 10^8$	$>6.0 \times 10^8$
32	$>2.4 \times 10^9$	$>2.6 \times 10^8$	$>6.0 \times 10^8$	$>6.0 \times 10^8$

In Table 6 the bacterial counts (raw data) are shown for individual mice, analysed five days after injection of *Listeria*.

G-CSF deficient mice do not have a readily mobilisable pool of neutrophils, as illustrated by the fact that their neutrophil counts are low 1 hour after *Listeria* infection compared to wild-type mice. Although their neutrophil levels do increase in absolute terms during the course of *Listeria* infection, this haematological response is markedly impaired compared to that of the wild-type controls, never attaining even the baseline level of the control group. This impaired haematological response correlates with the generally higher bacterial loads observed in the liver and spleen of these mice 5 days after infection.

<u>Example 16</u> Production of Mice Deficient In Both GM-CSF and CSF-1

GM-CSF deficient mice (genotype GM-/-) and appropriately outbred GM+/+ controls were generated as described above. Initial GM-/- M-/- were generated by interbreeding GM+/- M+/+ and GM+/+ M+/- mice (i.e. *op*/+) and selecting the GM+/- M+/- progeny for interbreeding. These matings generated the various genotypes in appropriate ratios, including GM-/- M-/- and GM+/+ M-/- (i.e. *op/op*). To generate additional GM-/- M-/- animals, GM-/-M+/- mice were mated.

<u>Example 17</u> Viability and Fertility of GM-/- M-/- Mice

From initial matings of GM+/- M+/- x GM+/- M+/- mice, litters of 7±2 pups (n=15) were born. The nine possible genotypes were represented in approximately the expected Mendelian ratios at weaning amongst 144 pups, as summarized in Table 8.

Table 8

Ratio of Genotypes in Litters From Matings of Mice Heterozygous for Both M-CSF and GM-CSF Null Mutations

Genotype		Expected ratio		Observed	
GM	M	of 16	of 144	^{n}total	^{n}male(%)
+/+	+/+	1	9	10	5 (50)
+/+	+/-	2	18	16	6 (38)
+/+	-/-	1	9	10	6 (60)
+/-	+/+	2	18	19	11 (58)
+/-	+/-	4	36	28	16 (57)
+/-	-/-	2	18	15	9 (60)
-/-	+/+	1	9	17	11 (65)
-/-	+/-	2	18	21	6 (29)
-/-	-/-	1	9	8	6 (75)

Results from 144 genotyped progeny of 4 double-heterozygous breeding pairs. During this period, the following mice were not genotyped: one litter of 9 pups that died at 3 days, 2 GM?M-/- males, 2 GM-/-M? males, one GM?M? female.

In particular, GM-/- M-/- mice were not significantly under-represented relative to wild-type mice, indicating that GM-/- M-/- mice were viable and there was no major fetal or neonatal loss, although amongst this genotype, male mice were over-represented. Subsequently, GM-/- M+/- x GM-/- M+/- matings generated litters of which 25/164 (15%) of pups were GM-/- M-/-; this was significantly fewer than expected ($p<0.05$), and there was again a male preponderance (19/25) ($p<0.1$). GM-/- M-/- males were fertile; one such male mated with a GM-/- M+/-

female has sired 5 litters of 5±3 pups. We have not bred from M-/- females regardless of their GM-CSF status, in view of the difficulties experienced by others.

The Chi-squared test was used to test for differences amongst viability and fertility data. An unpaired t-test was used to test for statistically significant differences in hematological data. Kaplan-Meier survival curves were constructed to compare survival data, and median survivals were tested for difference using the Mantel-Cox statistic. For survival data, mice killed because of apparent illness were regarded as non-survivors. A p-value of ≤ 0.05 was regarded as statistically significant.

Survival of GM-/- M-/- was compared to that of all other toothless mice (GM+/- M-/- and GM+/+ M-/-), and mice deficient only in GM-CSF (GM-/- M+/+) and wild-type (GM+/+ M+/+) mice. All toothless (M-/-) mice had significantly reduced survival compared with wild-type and GM-CSF deficient littermates, regardless of whether they were only M-CSF deficient, or both GM- and M-CSF deficient (Figure 2). Median survivals were: GM+/+M+/+ and GM-/- M+/+, not attained with median follow-up for 7 months; M-/- GM+/+ or +/-, 231 days (n=27); GM-/- M-/-, 71 days (n=38). The median ages at death for these two groups of M-/- mice were 33 and 41 days respectively. The median survival of M-/- GM+/+ or +/- mice was significantly shorter than that of GM+/+ M+/+ mice (p=0.03), and that of GM-/- M-/- mice was significantly shorter than that of other M-/- mice (p=0.01). Animal husbandary difficulties contributed to the death of most M-/- mice, but at histological examination 19/19 GM-/- M-/- mice that died or were killed at ages from 24 days to 324 days because of apparent illness had severe lung disease and acute lobar pneumonia at the time of death; some of the mice killed at ages of 24 to 324 days because of apparent distress were tachypnoeic.

__Figure 18__ **Phenotypic Features of GM-/- M-/- Mice**

GM-/- M-/- mice showed the typical gross phenotypic features of *op/op* mice ie. osteopetrosis, failure of tooth eruption, and domed head shape. In addition, they all had diseased lungs with many of the features typical of the lung disease observed in GM-/- mice. The GM-/- M-/- mice have age-for-age a more severe and more persistent form of the pathology seen in GM-/- M+/+ mice.

__Example 19__ **PCR diagnosis of GM-CSF and M-CSF genotype**

The allelic status of mice at the GM-CSF locus was determined by PCR as described above. GM-CSF genotypes are designated as follows: wild-type, GM+/+, heterozygous [wild-type/disrupted] GM+/-; and homozygous for the disrupted allele, GM-/-.

Whilst *op/op* (M-CSF genotype designated M-/-, in parallel to GM-CSF genotypes given above) are readily recognized in litters from M+/- x M+/- matings by their "toothless" phenotype, heterozygous (M+/-) and wild-type (M+/+) littermates are not distinguishable by phenotypic features, and in experiments with *op/op* (M-/-) it has been typical to use a random selection of M+/- and M+/+ toothed littermates as controls without precise definition of their genotypic status (40,41). In order to eliminate this imprecision and the heterogeneity that a possible gene-dosage effect might introduce, particularly in the face of GM-CSF deficiency, a PCR-based strategy for defining allelic status at the M-CSF locus was devised. By comparison of the murine cDNA sequence (42,43) with the published incomplete genomic structure of the human M-CSF gene (44), it was predicted that the *op* point mutation (45) lay in an exon homologous to human exon 3. Primers complementary to murine cDNA sequences were used to amplifiy a fragment from murine genomic DNA corresponding to the intron 5' of this exon, and 300 bp of the 3' end of this 1.6 kb PCR-generated fragment sequenced. The

diagnostic PCR paired a 5' sense primer from this intronic sequence (5'-TGTGTCCCTTCCTCAGATTACA-3') with a 3' antisense primer (5'-GGTCTCATCTATTATGTCTTGTACCAGCCAAAA-3') to generate a PCR product of 195 bp or 196 bp dependent on whether or not the thymidine base insertion of the *op* mutation was present in template DNA. Consequent to a 2 bp mis-match in the 3'-antisense primer (underlined), this PCR product contained a second *Bgl*I site dependent on whether template contained the *op* thymidine insertion. The PCR product was digested with *Bgl*I which cut at the intronic *Bgl*I site generating a 96 bp product (confirming restriction enzyme activity in each PCR reaction), and also, in the case of product generated from *op* template, cut the resulting 100 bp *op*-template-dependent fragment into diagnostic 70 and 30 bp fragments, but not the 99 bp wild-type-dependent fragment. Heterozygosity (*op*/+, M+/-) could be distinguished from homozygosity (*op/op*, M-/-) on the basis of phenotype, but also on the basis of whether the fragment of 100 bp had been completely digested by *Bgl*I, or whether uncut 99 bp fragment remained.

Ethidium bromide stained agarose gel of PCR products generated from wild-type, *op/op* and *op*/+ genomic DNA, demonstrated the diagnostic 70 bp and 30 bp fragments in *Bgl*I digests of PCR products from *op*-containing template DNA. The latter fragment was visible only in the case of the *op/op* template.

Example 20 **Haematological Analysis of GM-/- M-/- Mice**

Haemoglobin and platelet levels in mice of the four genotypes mice did not differ significantly at 4-5 weeks of age, but by 18-24 weeks of age, GM-/-M-/- mice had significantly elevated hemoglobin levels. These results are summarized in Table 8. Markedly increased hemoglobin levels were found in some older GM-/- M-/- mice: e.g. one animal that was killed because of apparent distress had a hemoglobin level of 220 g/l.

Table 9

Peripheral Blood of 18-24 wk GM-CSF and M-CSF Deficient Mice

Genotype			Peripheral blood parameter						
			Hemoglobin	Platelets	Leukocytes	Neutrophils	Lymphocytes	Monocytes	Eosinophils
GM	M	n	g/l	$x10^9/l$	$x10^9/l$	$x10^9/l$	$x10^6/l$	$x10^9/l$	$x10^6/l$
+/+	+/+	4	174±4	741±138	9.3±2.3	1.6±0.6	7.1±1.6	0.21±0.09	0.35±0.26
-/-	+/+	5	174±5	929±183	8.1±2.3	1.3±0.5	6.5±1.7	0.22±0.11	0.12±0.12
+/+	-/-	3	172±4	823±65	6.7±1.4	1.2±0.2	5.2±1.6	0.15±0.06	0.08±0.10
-/-	-/-	7	185±10*#⊗	604±70#⊗	6.6±1.1*	2.1±0.9#⊗	4.3±0.9*#	0.05±0.06*#	0.07±0.08*

n = number of mice

* = $p<0.05$ compared to GM+/+ M+/+

= $p<0.05$ compared to GM-/- M+/+

⊗ = $p<0.05$ compared to GM+/+ M-/-

GM-/- M-/- mice of 18-24 weeks of age had significantly reduced numbers of total leukocytes due primarily to a relative lymphopenia, as has been previously reported for M-/- mice (Table 2 and 2). Monocytes were present in the peripheral blood of GM-/- M-/- mice, but in mice of this age levels were significantly lower than in wild-type and GM-/- M+/+ mice and than those in GM+/+ M-/- mice (Table 2). GM-/- M-/- had significantly more circulating neutrophils than mice deficient in either factor alone, although not significantly more than their age-matched factor-replete controls. This relative neutrophilia persisted throughout life: GM-/- M-/- mice of 4-5 weeks of age had granulocyte levels of $3.0 \pm 2.4 \times 10^9$/L (n=4) compared to $1.3 \pm 0.7 \times 10^9$/L for age-matched GM+/+ M+/+ (n=33) and 3 older mice of age 26, 31 and 47 wks had granulocyte levels of 1.74, 4.37 and 7.7 $\times 10^9$/l respectively. Despite the relative monocytopenia, the diseased lungs of GM-/- M-/- mice contained large numbers of foamy alveolar macrophages, indicating that at least qualitatively monocytes were available for host defence in tissues.

Mice at 4-5 weeks, 18-24 and 28 weeks, were also compared at the results <u>are</u> summarized in Table 9. These results are consistent with those in Table 8.

<u>Table 10</u>

Peripheral Blood of GM-CSF and M-CSF Deficient Mice

Age	Genotype			Peripheral blood parameter						
Age	Genotype			Hemoglobin	Platelets	Leukocytes	Neutrophils	Lymphocytes	Monocytes	Eosinophils
wks	GM	M	n	g/l	$x10^9$/l	$x10^9$/l	$x10^9$/l	$x10^9$/l	$x10^9$/l	$x10^9$/l
4-5	+/+	+/+	16	134±18	771±86	6.9±2.8	1.3±0.7∇	5.6±2.4	0.03±0.04	ND
	-/-	+/+	33	142±20	724±216	6.8±2.6	1.7±1.5∇	5.1±2.0	0.07±0.16	ND
	+/+	-/-	nil						0.17±0.09	
	-/-	-/-	4	147±5	747±131	5.8±2.7	3.0±2.4∇	2.7±2.4	0.17±0.09	ND
18-24	+/+	+/+	4	174±4	741±138	9.3±2.3	1.6±0.6	7.1±1.6	0.21±0.09	0.35±0.26
	-/-	+/+	5	174±5	929±183	8.1±2.3	1.3±0.5	6.5±1.7	0.22±0.11	0.12±0.12
	+/+	-/-	3	172±4	823±65	6.7±1.4	1.2±0.2	5.2±1.6	0.15±0.06	0.08±0.10
	-/-	-/-	7	185±10*#⊗	604±70#⊗	6.6±1.1*	2.1±0.9#⊗	4.3±0.9*#	0.05±0.06*#	0.07±0.08*
28	-/-	-/-	2	175	638	5.3	3.06∇	2.2	0.06	

n = number of mice
* = $p<0.05$ compared to GM+/+ M+/+
\# = $p<0.05$ compared to GM-/- M+/+
⊗ = $p<0.05$ compared to GM+/+ M-/-
∇ = total granulocyte count (ie. neutrophils and eosinophils)
ND = not determined

The mice of genotype GM-/-, M-/- show features typical of both the CSF-1-/- genotype, such as osteopetrosis and failure of tooth eruption, and of the GM-CSF-/- genotype, such as propensity to infection, especially pneumonia caused by Gram-negative organisms. These pneumonias are usually fatal. For example, in one group of four GM-/-, M-/- animals, one was found dead at the age of 8 weeks, and pneumonia caused by *Klebsiella oxytoca* was confirmed bacteriologically. Three animals were killed after being found seriously ill; one, killed at six weeks had *Klebsiella oxytoca* pneumonia, one, killed at six weeks had a facial abscess caused by a *Streptococcus* species, and one, killed at twenty weeks had pneumonia caused by *Pasteurella pneumotropica*.

These doubly deficient mice have lung pathology presenting the same features as seen in GM-CSF deficient mice. Despite the deficiency of CSF-1, however, these mice still have monocytes/macrophages at normal or near-normal levels. Thus notwithstanding the double-deficiency these mice are able to maintain steady-state granulopoiesis and monopoiesis.

References cited herein are listed on the following pages, and are incorporated herein by this reference.

The following words used herein are trade marks: Araldite, Epon, Genescreen Plus, Geneticin, Sysmex.

It will be apparent to the person skilled in the art that while the invention has been described in some detail for the purposes of clarity and understanding, various modifications and alterations to the embodiments and methods described herein may be made without departing from the scope of the inventive concept disclosed in this specification.

REFERENCES

1. Gasson J.C.
Blood, 1991 77 1131-1145.

2. Metcalf D., Begley C.G., Williamson D.J., Nice E.C., De Larmarter J.D., Mermod J-J., Thatcher D., Schmidt A.
Exp. Hematol., 1987 15 1-9.

3. Lang R.A., Metcalf D., Cuthbertson R.A., Lyons I., Stanley E., Kelso A., Kannourakis G., Williamson D.J., Klintworth G.K., Gonda T.J., Dunn A.R.
Cell, 1987 51 675-686.

4. Johnson G.R., Gonda T.J., Metcalf D., Hariharan I.K., Cory S.A.
EMBO J., 1989 8 441-448.

5. Metcalf D.
Science, 1991 254 529-533.

6. Mortensen B.T., Schifter S., Pedersen L.B., Jensen A.N., Hovgaard D., Nissen N.I.
Exp. Hematol., 1993 21 1366-1370.

7. Lieschke, G.J., Dunn, A.R.
"Physiologic Role of Granulocyte Colony Stimulating Factor: Insights from *In Vivo* Studies"
Molecular Biology of Haematopoiesis, 1992 2 201-216 (Intercept Ltd, Andover, U.K.)

8. Evans, M.J.
Mol. Biol. Med., 1989 6 557-565

9. Capecchi,
Trends in Genetics, 1989 5 70-76

10. Smithies, O.
Trends in Genetics, 1993 9 112-116

11. Koller, B.H. and Smithies, O.
Annual Rev. Immunol., 1992 10 705-

12. Corcoran, L.M.
Today's Life Science, 1994 6 14-20

13. Miyatake S., Otsuka T., Yokota T., Lee F., Arai K.
EMBO J., 1985 4 2561-2568.

14. Stanley E., Metcalf D., Sobieszczuk P., Gough N.M., Dunn A.R.
EMBO J., 1985 4 2569-2573.

15. Tybulewicz V.L.J., Crawford C.E., Jackson P.K., Bronson R.T., Mulligan R.C.
Cell, 1991 65 1153-1163.

16. Gough N.M., Metcalf D., Gough J., Grail D., Dunn A.R.
EMBO J., 1985 4 645-653.

17. Handyside A.H., O'Neil G.T., Jones M., Hooper M.L.
Roux's Arch. Dev. Biol., 1989 198 8-55.

18. Mann G.B., Fowler K.J., Gabriel A., Nice E.C., Williams R.L., Dunn A.R.
Cell, 1993 73 249-261.

19. Metcalf D.
The Hemopoietic Growth Factors (Elsevier, Amsterdam) 1984

20. Kelso A., Gough N.M.
Growth Factors, 1989 1 165-177.

21. Metcalf D.
Proc. Natl. Acad. Sci. USA, 1991 88 11310-11314.

22. Harlow E., Lane D.
Antibodies; a laboratory manual (Cold Spring Harbor Laboratory, NY) 1988

23. Morse H.C., Davidson W.F., Yetter R.A., Coffman R.L.
Cell. Immunol., 1982 70 311-320.

24. Yelton D.E., Desaymard C., Scharff M.
Hybridoma, 1981 1 5-11.

25. Dialynas D.P., Quan Z.S., Wall K.A., Pierres A., Quintans J, Loken M.R., Pierres M., Fitch F.W.
J. Immunol., 1983 131 2445-2451.

26. Ledbetter J.A., Herzenberg L.A.
Immunol. Rev., 1989 49 63-90.

27. Metcalf D.
Blood, 1993 82 3515-3523.

28. Shelburne J.D., Wisseman C.L., Broda K.R., Roggli V.L., Ingram P.
in Diagnostic Electron Microscopy, eds. Trump B.F. & Jones R.T. (John Wiley & Sons, New York), 1983 4 475-532.

29. Rosen S.H., Castleman B., Liebow A.A.
N. Engl. J. Med., 1958 258 1123-1142.

30. Nogee L.M., De Mello D.E., Dehner L.P., Colten H.R.
N. Engl. J. Med., 1993 328 406-410.

31. Jones C.C.
Am. J. Med., 1960 29 713-722.

32. Beck J.M., Warnock M.L., Curtis J.L., Sniezek M.J., Arraj-Peffer S.M., Kaltreider H.B., Shellito J.E.
Am. J. Respir. Cell Mol. Biol., 1991 5 186-197.

33. Bilyk N., Holt P.G.
J. Exp. Med., 1993 177 1773-1777.

34. Smith S.M., Lee D.K.P., Lacy J., Coleman D.L.
Am. J. Respir. Cell Mol. Biol., 1990 2 59-68.

35. Chrane, H.T., Metcalf, D., Cheers, C.
Immunology, 1990 71 377-382

36. Frenck, R.W., Sarman, Harper, T.E., Buescher, E.S.
J. Infect. Dis., 1990 162 109-114

37. Morrissey, P.J., Charrier, K.
J. Immunol., 1990 144 557-561

38. Iida, J., Saiki, L., Ishihara, C., Azuma,
Vaccine, 1989 7 229-233

39. Freund, M., Kleine, H.-D.
Infection, 1992 20 Suppl. 2 S88-S90

40. Marks, S.C., Lane, P.W.
J. Heredity, 1976 67 11-18

41. Wiktor-Jedrzejczak, W., Ahmed, A., Szczylik, C., Skelly, R.R.
J. Exp. Med., 1982 156 1516-1527

42. Ladner, M.B., Martin, G.A., Noble, J.A., Wittman, V.P., Warren M.K., McGrogan, M., Stanley, E.R.
Proc. Natl. Acad. Sci., U.S.A., 1988 85 6706-6710

43. DeLamarter, J.F., Hession, C., Semon, D., Gough, N.M., Rothenbuhler, R., Mermod, J-J.
Nuc. Acids Res., 1987 15 2389-90

44. Kawasaki, E.S., Ladner, M.B.
In Dexter TM, Garland JM Testa NG eds
Colony Stimulating Factors
New York: Marcel Dekker 1990 155-176

45. Yoshida, H., Hayashi, S-I.,Kunisada, T., Ogawa, M., Nishikawa, S., Okamura, H., Sudo, T., Shultz, L.D., Nishikawa, S-I.
Nature, 1990 345 442-444

46. Stanley, E., Leischke, G.J., Grail, D. *et al*
Proc. Natl. Acad. Sci., U.S.A., (in press)

47. Pollard, J.W., Hunt J.S., Wiktor-Jedrzejczak, W., Stanley, E.R.
Developmental Biol., 1991 148 273-283

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. A non-human animal carrying a disruption of a gene encoding a colony stimulating factor.

2. A non-human animal according to Claim 1 carrying a disruption of the gene encoding granulocyte-macrophage colony-stimulating factor (GM-CSF), such that expression of GM-CSF is disrupted.

3. A non-human animal according to Claim 2, in which said disruption results in inability of spleen cells from the non-human animal to produce detectable levels of bioactive GM-CSF.

4. A non-human animal according to Claim 2 or Claim 3, which is a mouse.

5. A non-human animal according to any one of Claims 2 to 4, in which the gene encoding GM-CSF is completely inactivated.

6. A non-human animal according to any one of Claims 2 to 5, which carries a mutation comprising deletion of exons 1 and 2 and intron 1 between *Sca*I and *Sma*I restriction sites of the gene encoding GM-CSF.

7. A non-human animal according to any one of Claims 2 to 6, which carries one or more additional mutations which result in disruption of a specific gene.

8. A non-human animal according to Claim 7, wherein said additional mutation causes disruption of a gene encoding a growth factor selected from the group consisting of granulocyte colony-stimulating-factor (G-CSF), colony-stimulating-factor-1 (CSF-1); leukaemia inhibitory factor (LIF), and transforming growth factor-β1 (TGF-β1), or of a cytokine.

9. A non-human animal according to Claim 8, wherein the cytokine is an interleukin.

10. A non-human animal according to Claim 9, wherein the interleukin is selected from the group consisting of interleukin-2, interleukin-3 and interleukin-6.

11. A non-human animal according to Claim 10, which is a mouse, in which the genes for GM-CSF and CSF-1 are

disrupted, and in which said mouse is produced by crossing a mouse according to any one of Claims 2 to 6 with an *op/op* mouse.

12. A non-human animal according to Claim 1 carrying a disruption of the gene encoding granulocyte colony-stimulating factor (G-CSF), such that expression of G-CSF is disrupted.

13. A non-human animal according to Claim 12, in which said disruption results in inability of lung cells from the animal to produce detectable levels of bioactive G-CSF.

14. A non-human animal according to Claim 12 or Claim 13, which is a mouse.

15. A non-human animal according to any one of Claims 12 to Claim 14, in which the gene encoding G-CSF is completely inactivated.

16. A non-human animal according to any one of Claims 12 to 15, which carries a mutation comprising a deletion from the *Nco*I restriction site in exon 1 which encompasses the translation initiation codon ATG to the *Bam*HI restriction site in exon 3 of the gene encoding G-CSF.

17. A non-human animal according to any one of Claims 12 to 16, which carries one or more additional mutations which result in disruption of a specific gene.

18. A non-human animal according to Claim 17, wherein said additional mutation causes disruption of a gene encoding a growth factor selected from the group consisting of granulocyte-macrophage colony stimulating-factor (GM-CSF), colony-stimulating-factor-1 (CSF-1); leukaemia inhibitory factor (LIF), and transforming growth factor-β1 (TGF-β1), or of a cytokine.

19. A non-human animal according to Claim 18, wherein the cytokine is an interleukin.

20. A method of testing the efficacy of a putative method of treatment of alveolar proteinosis, comprising the step of subjecting a non-human animal according to any one of Claims 2 to 11 to said putative treatment.

21. A method according to Claim 20, wherein said putative treatment is a chemotherapeutic agent.

22. A method according to Claim 20 or Claim 21, in which the putative treatment is selected from the group consisting of granulocyte-macrophage colony-stimulating factor (GM-CSF) granulocyte-colony stimulating factor (G-CSF), colony-stimulating-factor-1 (CSF-1); leukaemia inhibitory factor (LIF), and transforming growth factor-β1 (TGF-β1), or a cytokine.

23. A method according to Claim 21, wherein the chemotherapeutic agent is GM-CSF.

24. A method of testing the efficacy of a putative treatment for infectious disease, comprising the step of subjecting a non-human animal according to any one of Claims 1 to 19 to said putative treatment.

25. A method according to Claim 24, wherein said infectious disease is a bacterial, fungal or viral infection of the lung.

26. A method according to Claim 24 or Claim 25, in which said bacterial infection is an opportunistic infection of the type to which immunocompromised or immunodeficient individuals are prone, or an infection which is intractable to currently-available therapies.

27. A method according to Claim 26 wherein the opportunistic infection is one to which patients suffering from acquired immunodeficiency syndrome (AIDS) are susceptible.

28. A method of treatment of alveolar proteinosis and/or of pulmonary infection, comprising the step of administering to a subject in need of such treatment an effective amount of GM-CSF.

29. A method according to Claim 28, wherein GM-CSF is administered locally to the lung.

30. A method according to Claim 28 or Claim 29, in which said GM-CSF is administered as an aerosol, via a nebuliser or via an endotracheal tube.

31. A method according to any one of Claims 28 to 30,

wherein the GM-CSF is administered as adjunctive therapy to other treatments.

32. A method according to any one of Claims 29 to 31, wherein the dose of GM-CSF is from 0.1 μg/kg body weight to 20 μg/kg body weight.

33. A method of treatment of pulmonary infection, comprising the step of administering to a subject in need of such treatment an effective amount of G-CSF.

34. A method according to Claim 33, wherein G-CSF is administered locally to the lungs.

35. A method according to Claim 33 or Claim 34, in which said G-CSF is administered as an aerosol, via a nebuliser or via an endotracheal tube.

36. A method according to any one of Claims 33 to 35, wherein the G-CSF is administered as adjunctive therapy to other treatments.

37. A method according to any one of Claims 33 to 35, wherein the dose of G-CSF is from 0.1 μg/kg body weight to 20 μg/kg body weight.

38. A factor which is involved in regulation of steady-state haematopoiesis, and which is present in animals in which expression of GM-CSF is disrupted.

39. A factor according to Claim 38, which is present in animals whose spleen cells are incapable of producing detectable levels of bioactive GM-CSF.

40. A factor according to Claim 38 or Claim 39, which is also present in animals in which expression of both GM-CSF and G-CSF is disrupted.

41. A factor according to Claim 38 or Claim 39, which is also present in animals in which expression of both GM-CSF and CSF-1 is disrupted.

42. A method of treatment of a patient suffering from alveolar proteinosis, comprising the step of administering the gene encoding GM-CSF to said patient via somatic gene therapy .

43. A method according to Claim 42, in which the somatic gene therapy comprises the step of intravenous

administration of a liposomal formulation of cDNA encoding GM-CSF.

44. A method of diagnosis of a colony-stimulating factor deficiency, comprising the step of testing a tissue or cell sample from a subjct suspected of suffering from such a deficiency for the absence of the gene encoding said factor.

45. A method of diagnosis of GM-CSF deficiency or of G-CSF deficiency, comprising the step of testing cells or tissue from a subject suspected of suffering such deficiency for the presence of a gene encoding GM-CSF or G-CSF respectively.

46. A method according to Claim 44 or Claim 45, wherein the test is carried out using peripheral blood cells.

47. A method according to Claim 44 or Claim 45, wherein the test is carrying out using tissue obtained by lung biopsy.

48. A targeting vector for disruption of a gene encoding GM-CSF, as described herein with reference to Figure 1.

49. A targeting vector for disruption of a gene encoding G-CSF, as described herein with reference to Figure 5.

50. A cell line derived from a non-human animal according to any one of Claims 1 to 19.

51. A cell line according to Claim 50, which is a bone-marrow stromal cell line.

52. A composition suitable for local administration to the lung, comprising GM-CSF and/or G-CSF and a pharmaceutically-acceptable carrier.

LUDWIG INSTITUTE FOR CANCER RESEARCH

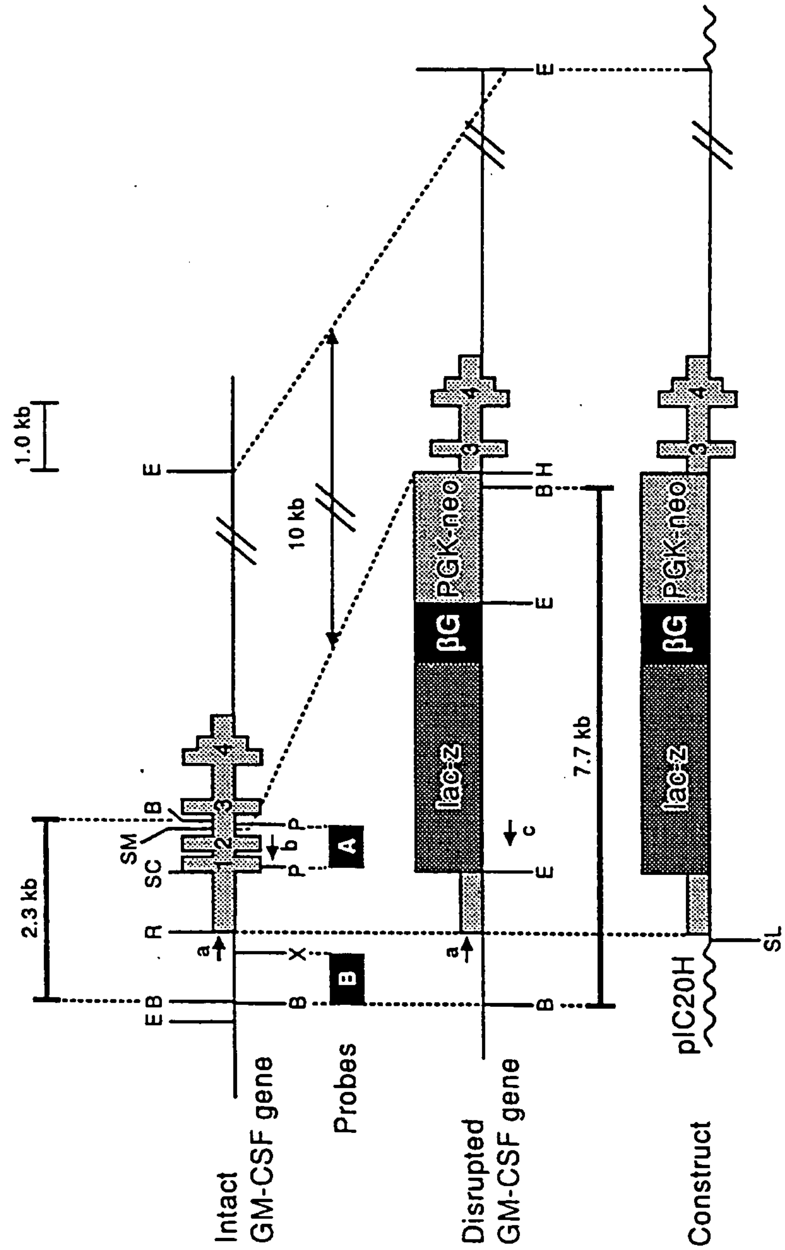

1.0 kb
2.3 kb
10 kb
7.7 kb
Intact GM-CSF gene
Probes
Disrupted GM-CSF gene
Construct
E
EB
R
SC
SM
B
P
X
A
B
a
b
c
H
βG
PGK-neo
lac-z
pIC20H
SL

FIGURE 1

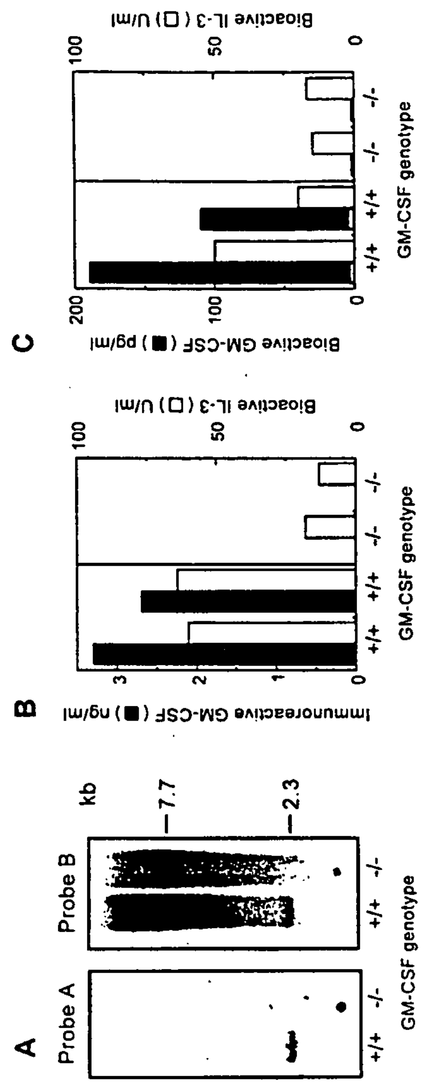

A
Probe A
Probe B
kb
7.7
2.3
+/+
-/-
GM-CSF genotype
B
Immunoreactive GM-CSF (■) ng/ml
Bioactive IL-3 (□) U/ml
C
Bioactive GM-CSF (■) pg/ml

FIGURE 2

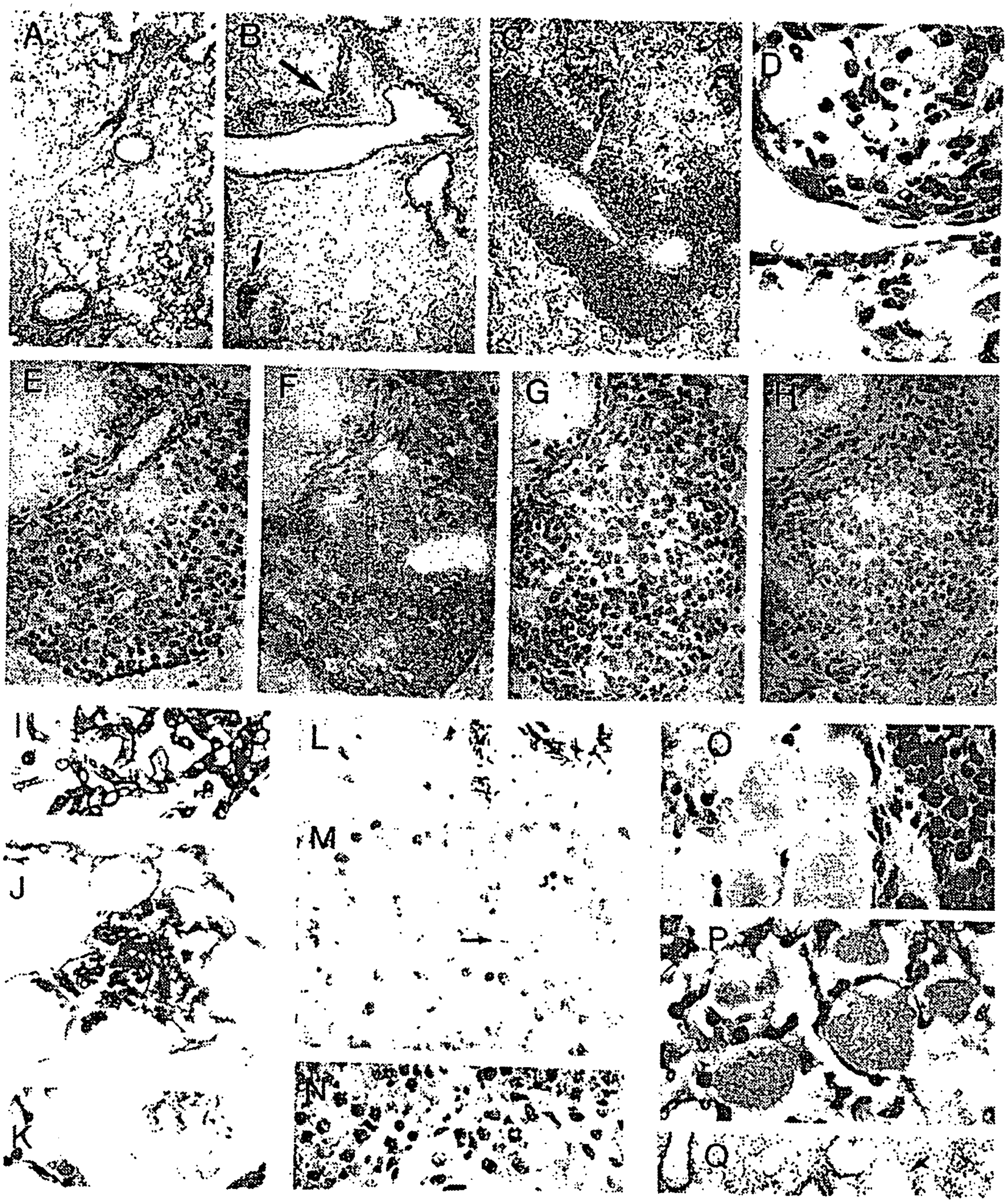


FIGURE 3

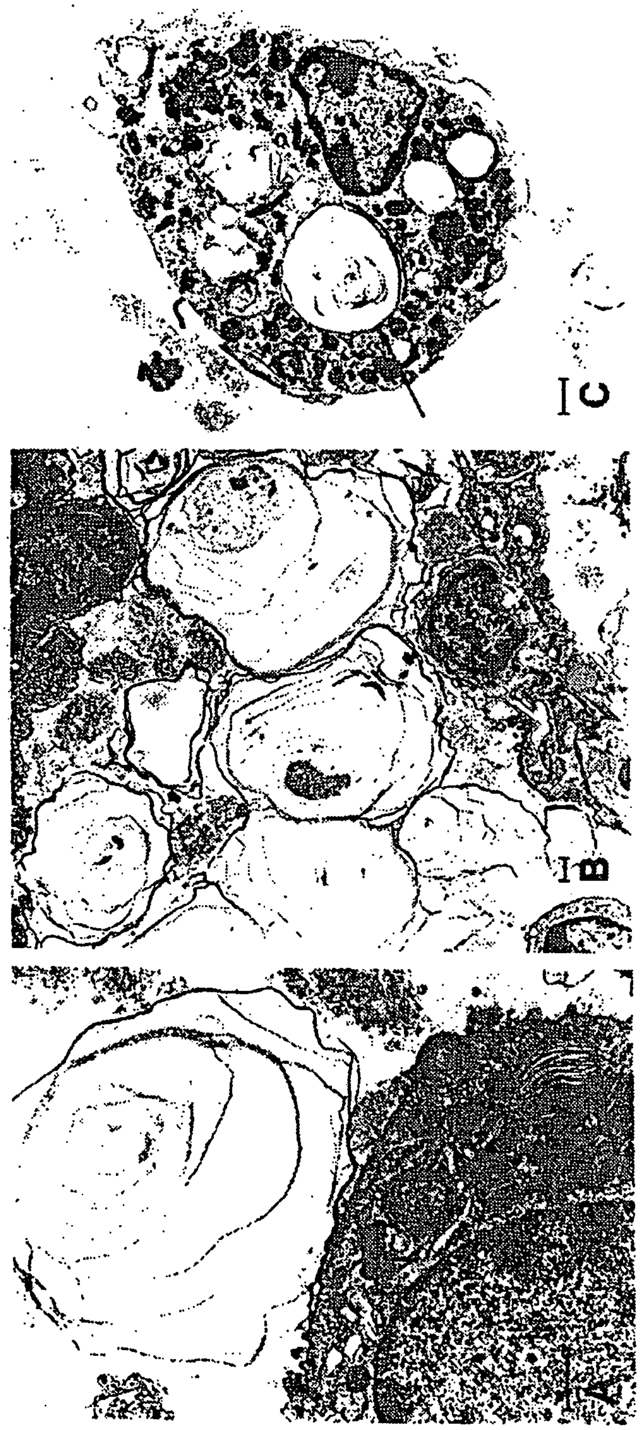


FIGURE 4

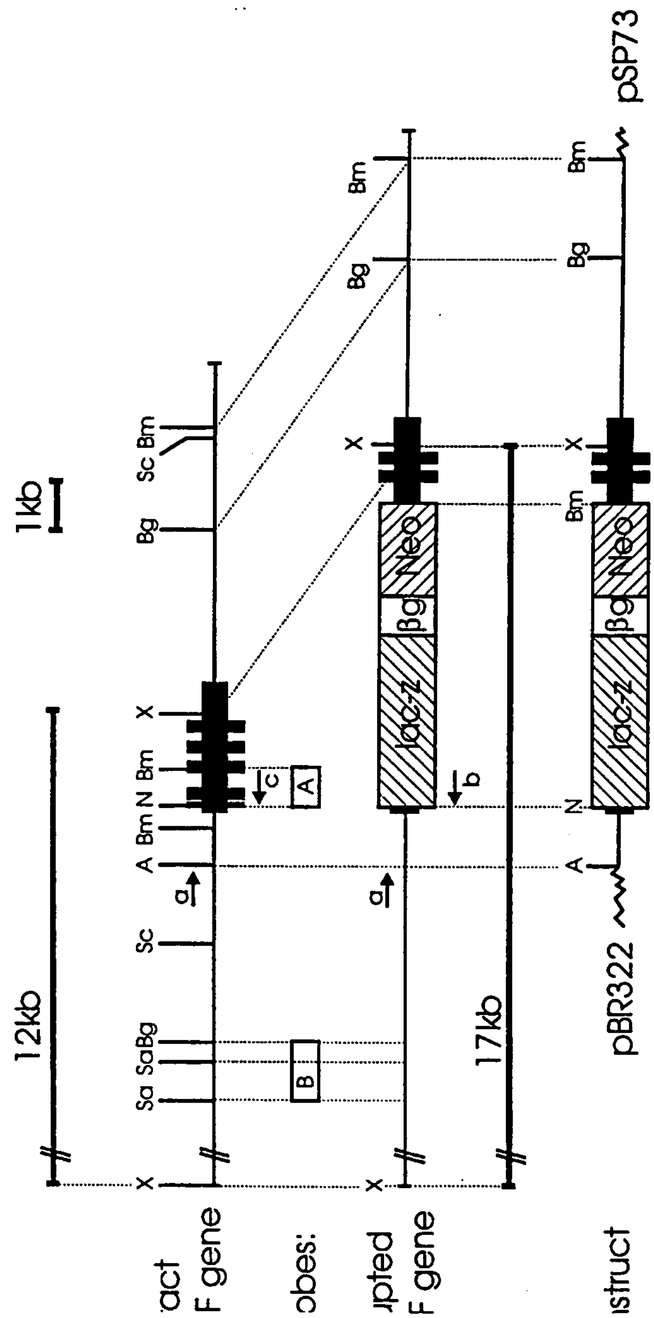

1kb
12kb
17kb
act
F gene
obes:
upted
F gene
istruct
pBR322
pSP73
lac-z
βg
Neo
Sa
SaBg
Sc
Bm
Bg
A
N
X
a
b
c
B

FIGURE 5a

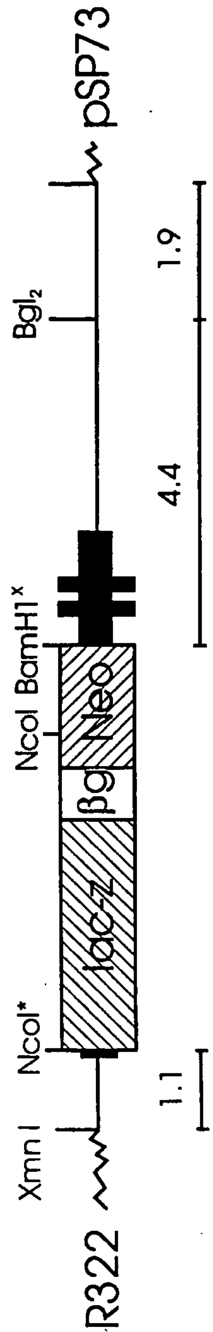


FIGURE 5b

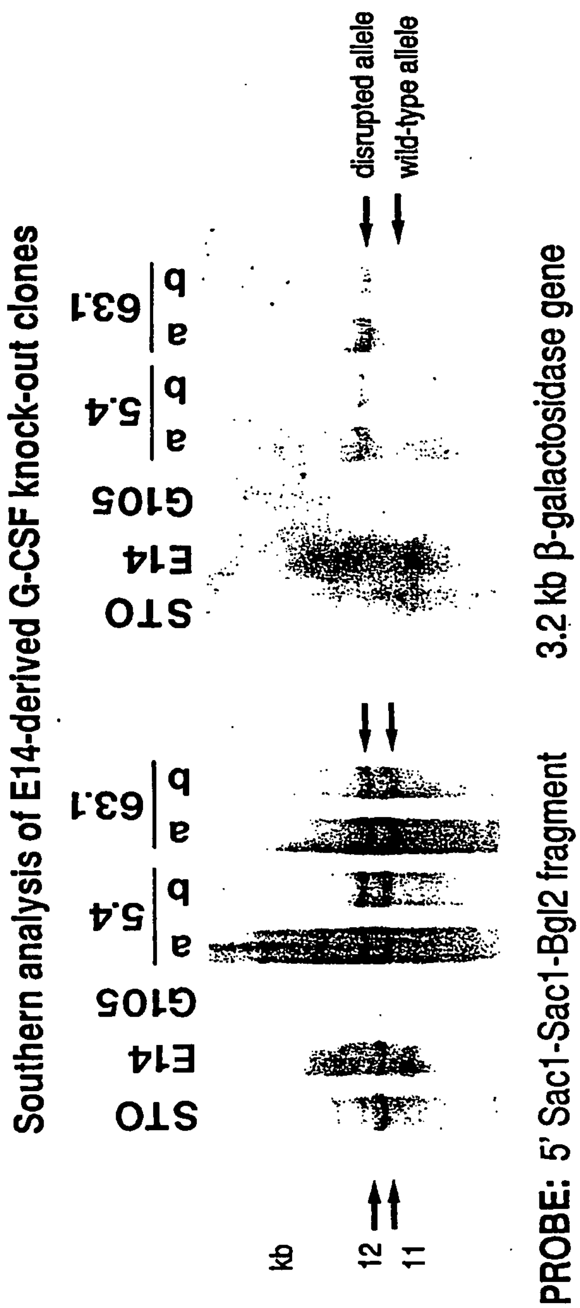

Southern analysis of E14-derived G-CSF knock-out clones
STO
E14
G105
5.4
a
b
63.1
a
b
disrupted allele
wild-type allele
kb
12
11
PROBE: 5' Sac1-Sac1-Bgl2 fragment
3.2 kb β-galactosidase gene

FIGURE 6

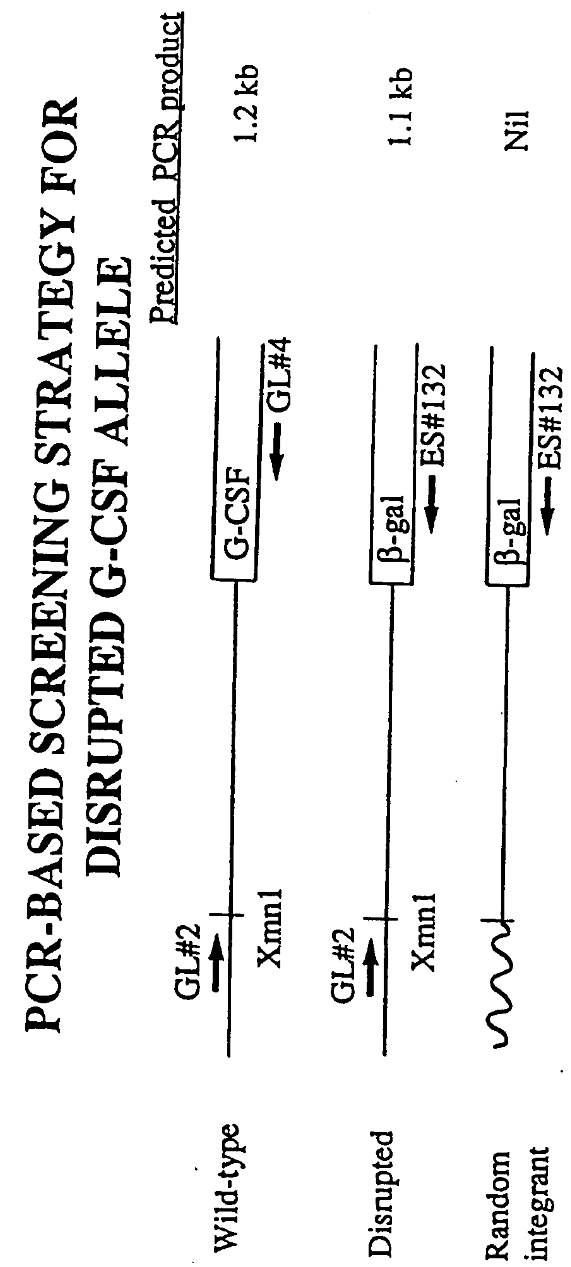

PCR-BASED SCREENING STRATEGY FOR
DISRUPTED G-CSF ALLELE
Predicted PCR product
Wild-type
GL#2
Xmn1
G-CSF
GL#4
1.2 kb
Disrupted
GL#2
Xmn1
β-gal
ES#132
1.1 kb
Random
integrant
β-gal
ES#132
Nil

FIGURE 7a

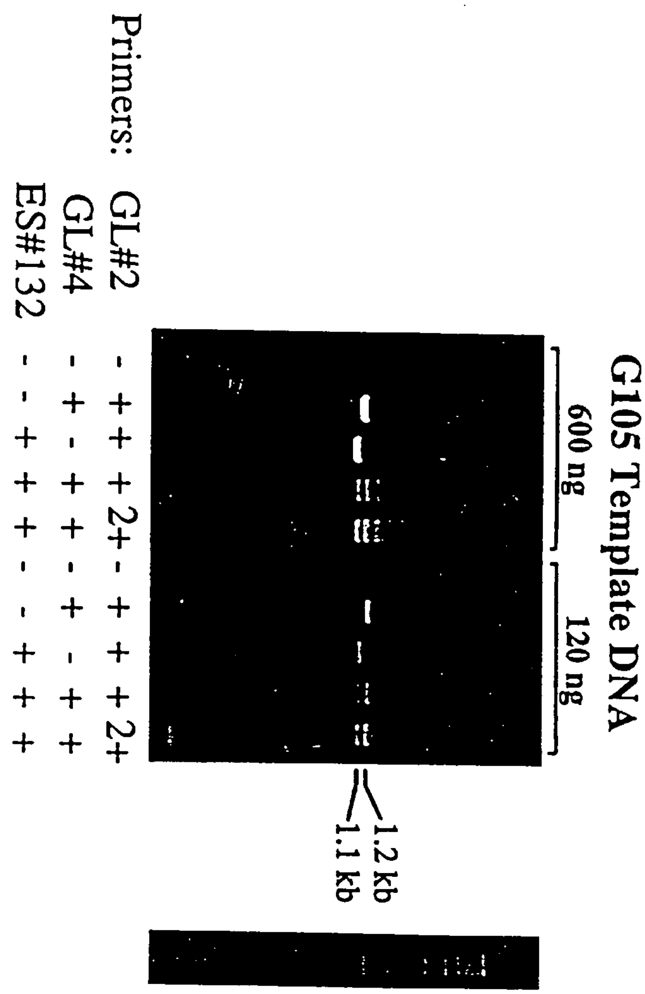

G105 Template DNA
600 ng
120 ng
1.2 kb
1.1 kb
Primers: GL#2 - + + + 2+ - + + + 2+
GL#4 - + - + + - + - + +
ES#132 - - + + + - - + + +

FIGURE 7b

FIRST LITTER FROM G-CSF chimera (≈+/-) x +/- MATING

PUPS: 1 2

FIGURE 9

PCR:

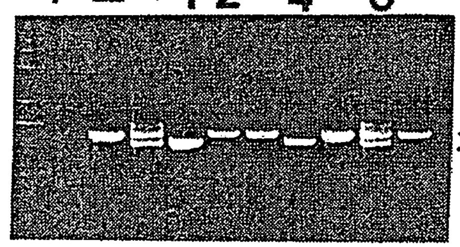


FIGURE 8

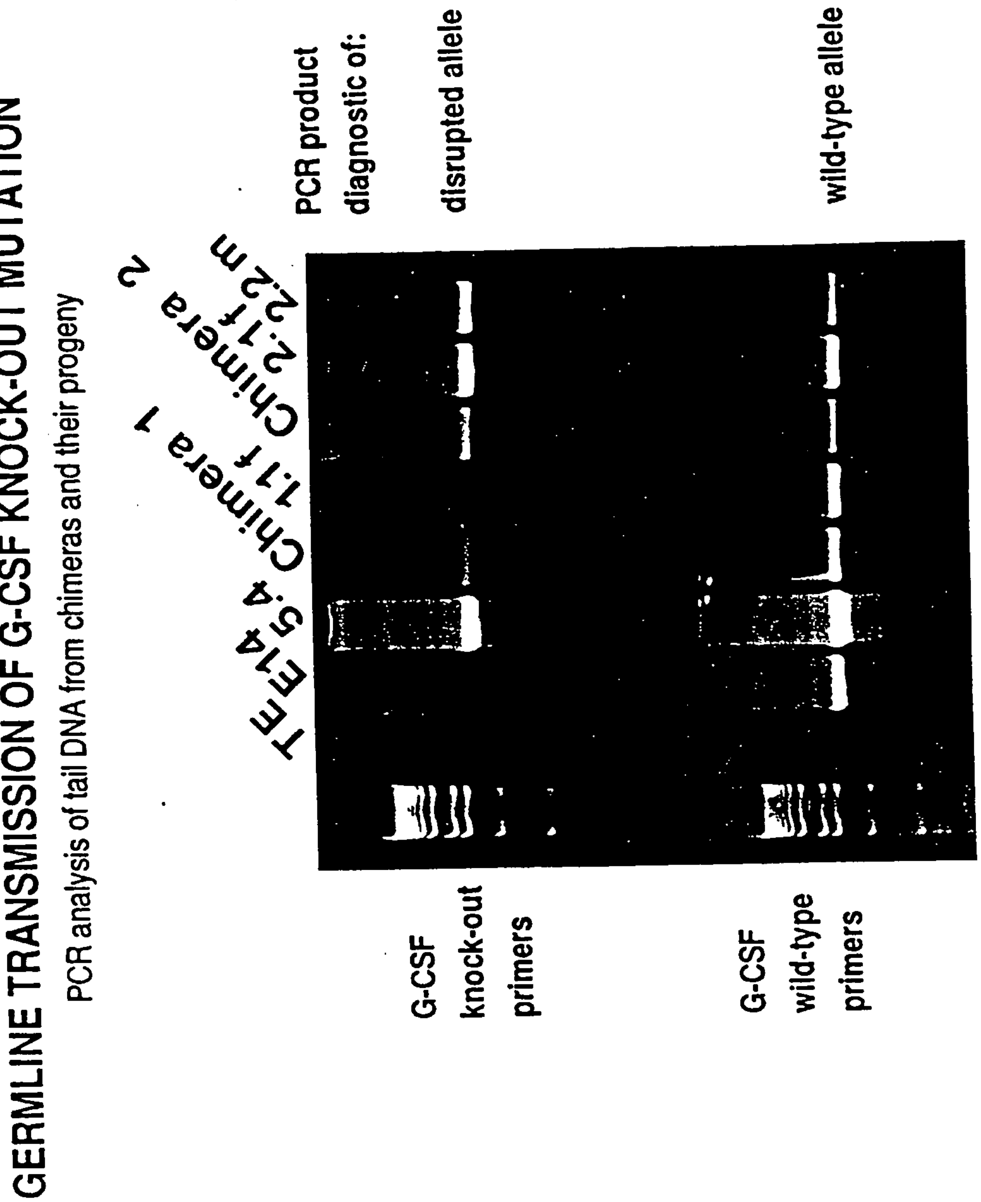

GERMLINE TRANSMISSION OF G-CSF KNOCK-OUT MUTATION
PCR analysis of tail DNA from chimeras and their progeny
PCR product diagnostic of:
disrupted allele
wild-type allele
Chimera 2
2.2 m
2.1 f
Chimera 1
1.1 f
5.4
TEE14
G-CSF knock-out primers
G-CSF wild-type primers

FIGURE 10

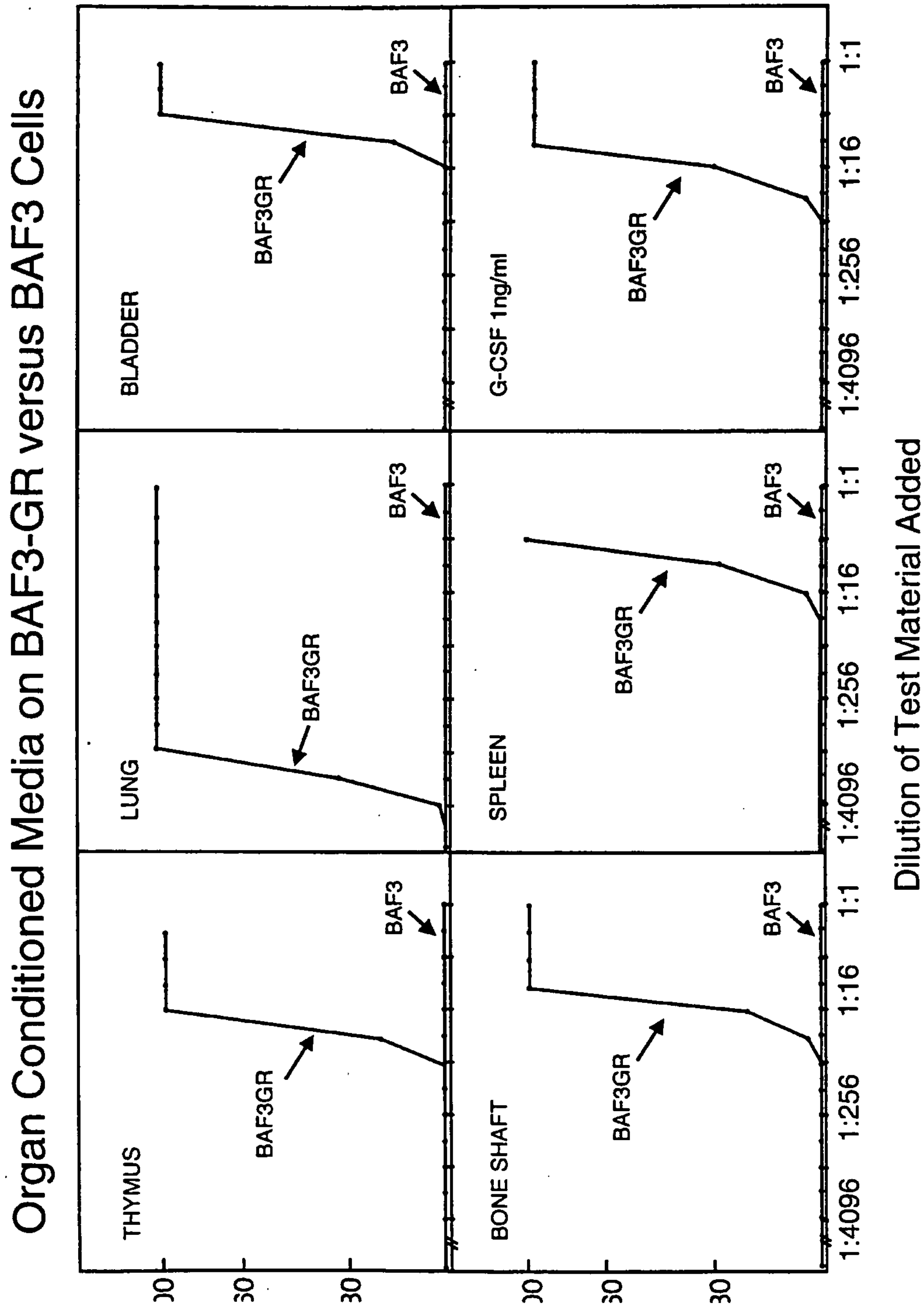

Organ Conditioned Media on BAF3-GR versus BAF3 Cells
THYMUS
LUNG
BLADDER
BONE SHAFT
SPLEEN
G-CSF 1ng/ml
BAF3GR
BAF3
1:4096
1:256
1:16
1:1
Dilution of Test Material Added

FIGURE 11

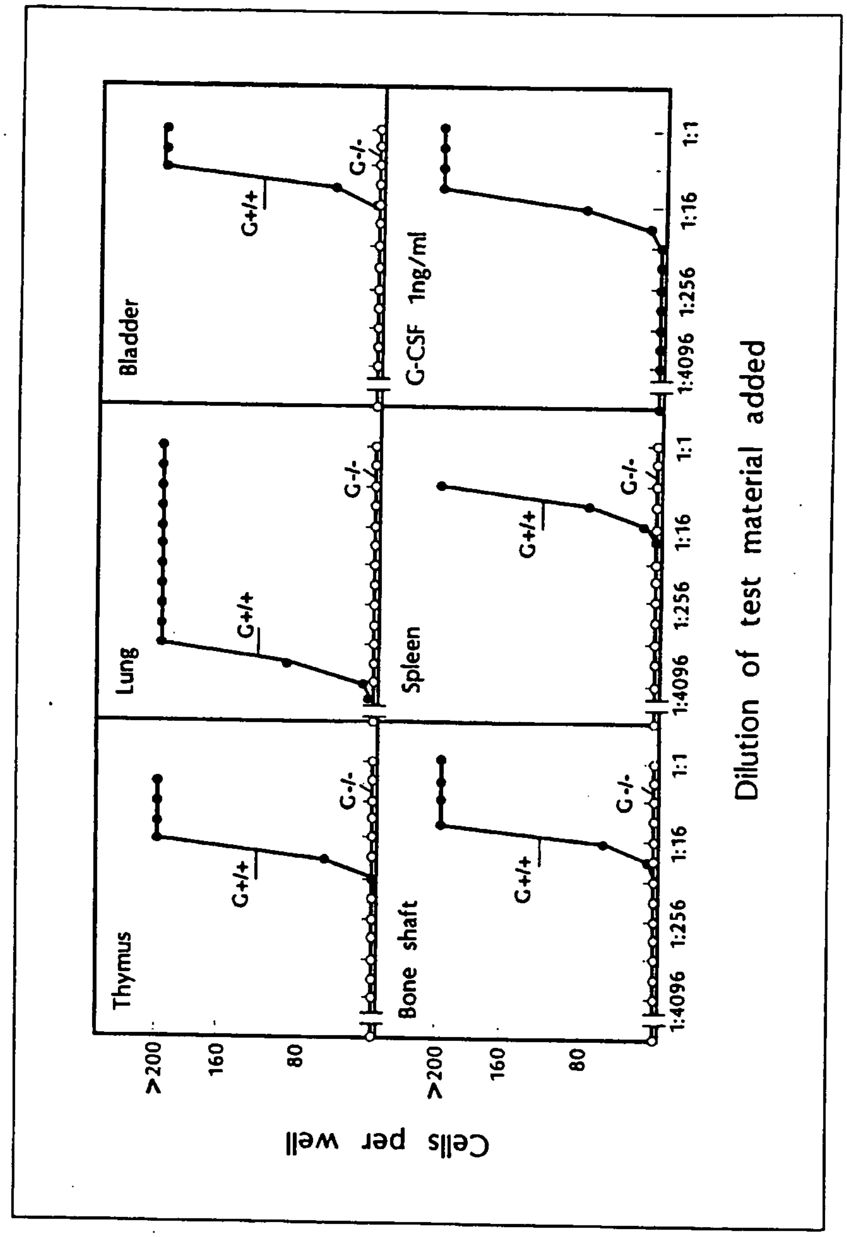

Thymus
Lung
Bladder
Bone shaft
Spleen
G-CSF 1ng/ml
G+/+
G-/-
>200
160
80
Cells per well
1:4096
1:256
1:16
1:1
Dilution of test material added

FIGURE 13a

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU 94/00103

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl.[5] C12N 015/27, A01K 067/027, C12Q 001/68, C07K 15/06, C07G 15/00, A61K 037/02

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
WPAT AND CHEM ABS
SEE DETAILS IN ELECTRONIC DATABASE BOX BELOW

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
AU A01K 067/027, C12N 015/27, A61K 037/02, A61K 048/00
BIOTECH (SEE DETAILS IN ELECTRONIC DATABASE BOX BELOW)

Electronic data base consulted during the international search (name of data base, and where practicable, search terms used)
DERWENT WPAT DATABASE, KEYWORDS: COLONY (W) STIMULATING (W) FACTOR # OR CSF AND DISRUPT: OR INSERT: OR DELET: OR MODIF: OR MUTAT OR MUTANT # OR MUTEIN #

CHEMICAL ABSTRACTS DATABASE, KEYWORDS AS ABOVE
BIOTECHNOLOGY ABSTRACTS DATABASE, KEYWORDS AS ABOVE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
X	EP, 0505123 A1 (AMGEN INC.) published 23 September 1992 (23.09.92) see whole document	33-37 and 52
X	EP, 0373679 (AMGEN INC.) published 20 June 1990 (20.06.90) see whole document	33-37 & 52

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

*	Special categories of cited documents :
"A"	document defining the general state of the art which is not considered to be of particular relevance
"E"	earlier document but published on or after the international filing date
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O"	document referring to an oral disclosure, use, exhibition or other means
"P"	document published prior to the international filing date but later than the priority date claimed
"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&"	document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
1 July 1994 (01.07.94)	4 July 1994 (04.07.94)

Name and mailing address of the ISA/AU	Authorized officer
AUSTRALIAN INDUSTRIAL PROPERTY ORGANISATION PO BOX 200 WODEN ACT 2606 AUSTRALIA	Arati Sardana ARATI SARDANA
Facsimile No. 06 2853929	Telephone No. (06) 2832627

Category*	Citation of document, with indication, where appropriate of the relevant passages	Relevant to Claim No.
X	AU 65493/86 (DE LAMARTER, JOHN) published 9 April 1987 (09.04.87) See claim 16	52
X	Australian Biotechnology, April 1993, 3, 2, pages 73-75 (Elwood and Begley): "Human Gene Therapy and the Clinical Implications" See whole article	42
X	Clinical Chemistry, (1994), 40, 4 pages 510-512. (K.W. Culver) "Clinical Applications of gene therapy for cancer" see whole article [note: the paper was presented at the San Diego conference: Beyond DNA probes, San Diego, CA, November 18-20, 1993]	42
P X	AU 48450/93 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) published 17 March 1994 (17.03.94) See whole document	42
Y	Annual. Rev. Immunol. (1992), 10, pages 705-30 (Koller and Smithies) "Altering genes in animals by gen targeting". See whole article	1-19 48-51
Y	Oncogene (1990), 5, pages 1799-1807 (C. Heberlein et al) "Retrotransposons as mutagens in the induction of growth autonomy in hematopoietic cells". See page 1805 column 2 paragraph 2	1-19 48-51
Y	Eur. J. Biochem. (1987), 169, pages 353-358. (N.J. Gough et al) "Mutagenesis of murine granulocyte/macrophage - colony - stimulating factor reveals critical residues near the N terminus". See abstract	1-19 48-51
A	Mol. Biol. Med. (1989), 6, pages 557-565 (M.J. Evans) "Potential for Genetic Manipulation of Mammals". See whole article.	1-19 48-51
A	Cold Spring Harbor Symposia Quant Biol, 51, (Mol. Biol. Homo Sapients, Pt. 2), 86.00.00 pages 899-909 (Le Beau et al) "Assignment of the GM-CSF, CSF-1, and FMS genes to Human Chromosome 5 Provides Evidence for linkage of a family of Genes Regulating Hematopoiesis and for their Involvement in the Deletion (5q) in Myloid Disorders". See page 904	1-19 48-51

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report		Patent Family Member					
EP	505123	AU	17476/92	AU,B2	643141	JP	5507944
		NZ	241954	US	5284656		
EP	373679	AU	46689/89	AU,B2	621695	JP	3502808
		US	5104651	WO	9006762		
AU	65493	DK	2828/87	EP	238655	WO	8702060
AU	48450	NONE					
							END OF ANNEX

Exhibit 8

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/00, 15/12, 15/10, 15/90, 5/10, C07K 14/47, A01K 67/027

A1

(11) International Publication Number: WO 99/09150

(43) International Publication Date: 25 February 1999 (25.02.99)

(21) International Application Number: PCT/US97/14507

(22) International Filing Date: 18 August 1997 (18.08.97)

(71) Applicant: BAYER CORPORATION [US/US]; 100 Bayer Road, Pittsburgh, PA 15205 (US).

(72) Inventor: WIRAK, Dana, Owen; 27 Barberry Lane, Woodbridge, CT 06525 (US).

(74) Agents: GREENMAN, Jeffrey, M. et al.; Bayer Corporation and Indiana Corporation, 400 Morgan Lane, West Haven, CT 06516-4175 (US).

(81) Designated States: AL, AM, AU, AZ, BA, BB, BG, BR, BY, CA, CN, CU, CZ, EE, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, RO, RU, SD, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

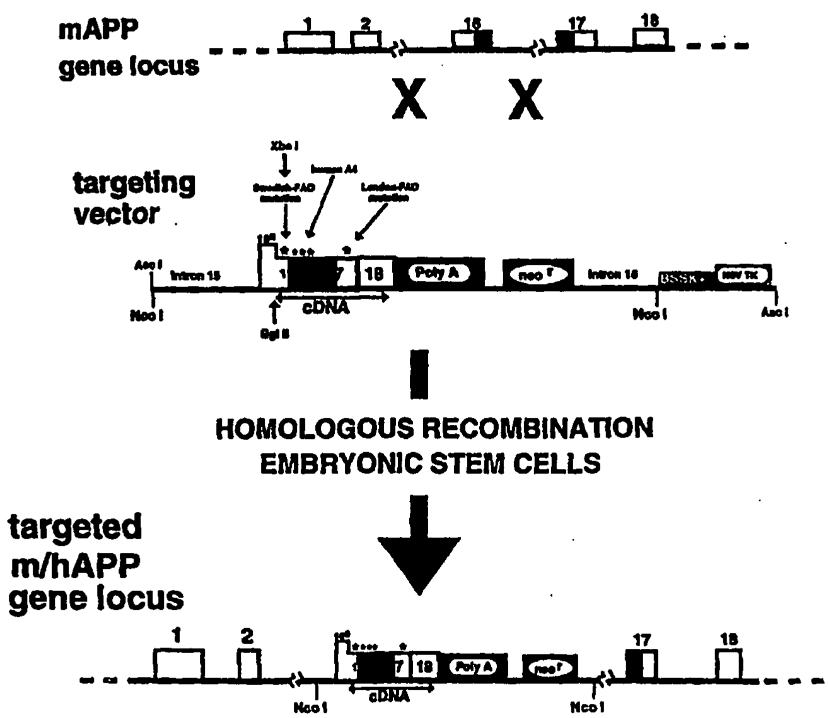
Published

With international search report.

(54) Title: METHOD OF INTRODUCING MODIFICATIONS INTO A GENE

(57) Abstract

The present invention relates to a gene targeting vector and a method of using it to modify nucleic acid sequences. A gene targeting vector in accordance with the invention can comprise: a nucleotide sequence which is effective to achieve homologous recombination at a predefined position of a target gene, operably linked to the 5' terminus of a nucleotide coding sequence which, when inserted into a target gene, codes for at least one amino acid whose identity and/or position is not naturally-occurring in the target gene, and a nucleotide sequence which is effective to achieve homologous recombination at a predefined position of the target gene, operably linked to the 3' terminus of said nucleotide coding sequence. The nucleotide coding sequence can code without interruption for an amino acid sequence, where the amino acid sequence is coded for by two or more exons in a naturally-occurring gene.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

METHOD OF INTRODUCING MODIFICATIONS INTO A GENE

BACKGROUND

The current prior art methods of modifying genes are cumbersome and difficult. For example, mutagenesis of the mouse gene locus via "hit-and-run" and "tag-and-exchange" gene targeting technologies can require the mouse gene locus to be targeted two times in succession using the same ES cell clone. This is a long and laborious process. It is extremely difficult to maintain totipotency of the ES cell through so many manipulations and over such long periods of time in culture. To overcome the difficulties in the prior art, we have developed a novel method of targeting and engineering gene sequences.

The current state of the art provides for three different approaches to the development of transgenic animal models (Lamb, *Nat. Genet.*, 9:4-6, 1995). The first approach utilizes pronuclear injections of recombinant minigenes into the pronuclei of 1-cell embryos. In the second approach, a complete gene residing in yeast artificial chromosomes (YACs), is electroporated into embryonic stem cells (ES cells). The third approach utilizes gene targeting techniques in ES cells to introduce point mutations into a gene present in the ES cell chromosome. The most common approaches to introducing point mutations are "hit-and-run" (Hasty et al., *Nature*, 350:243-246, 1991) or "tag-and-exchange" (Askew et al., *Mol. Cell. Biol.*, 13:4115-4124, 1993), (Stacey et al., *Mol. Cell. Biol.*, 14:1009-1016, 1994) gene targeting procedures.

Recombinant minigenes, when injected into mouse embryos, integrate into the mouse chromosome at random locations. The site of integration can often exert a deleterious influence on the pattern of expression and/or expression level of the recombinant level of the recombinant minigene ("position effect") (Bonnerot et al., *Proc. Natl. Acad. Sci.*, 87:6331-6315,

1990; Brinster et al., *Proc. Natl. Acad. Sci.*, 85:836-840, 1988; Grosveld et al., *Cell*, 51:976-985, 1987).

To illustrate the benefit and ease of the novel compositions, methods, treatments, etc., described herein, we have utilized genes associated with Alzheimer's disease. Therefore, although aspects of this disclosure are written with respect to Alzheimer's diseases, e.g.,the APP gene, it is recognized that this invention is in no way limited to such genes and diseases, but may be applied to any nucleic acids., etc., that one desires to target and/or modify.

Alzheimer's disease (AD) is a neurodegenerative disorder characterized by progressive deterioration of memory and cognition. Prominent histopathological features of this disease include the extracellular deposition of amyloid and the accumulation of intracellular neurofibrillary tangles. The principal underlying cellular features of AD are the degeneration affects many types of neurons and may account for the numerous neurological deficits that patients afflicted with the disease encounter. The most notable degeneration occurs in the hippocampus, cerebral cortex, and amygdala, regions of the brain that play a major role in memory, cognition, and behavior.

Although numerous attempts have been made to generate transgenic mouse models for AD via the pronuclear injection approach (Lamb, 1995), only one line of transgenic mice has succeeded in developing extra-cellular plaque-like deposits of beta-amyloid (Games et al., *Nature*, 373:523-527, 1995). This transgenic mouse line utilizes the PDGF promoter to over-express (> 10 fold) human "London"-FAD APP. Because of the artificial nature of the transgene's regulation of gene expression and the aberrant high levels of APP expression, the accumulation of amyloid in this line of transgenic mice may not be fully relevant to the cellular mechanisms involved in Alzheimer's disease.

Two additional papers report only partial success in developing AD-like pathology. In one transgenic model, human APP 751 is over-produced in the brain using the brain-specific enolase promoter (Higgins et al., *Ann. Neurol.*, 35:598-607, 1994). This mouse model exhibits diffuse extra-cellular staining for beta-amyloid, but there was no evidence of accumulations of plaque-like deposits as described by Games et al. (Games et al., 1995). Another transgenic model exhibits intra-cellular deposits of beta-amyloid (La Ferla et al., *Nat. Genet.*, 9:21-30, 1995). This deposition leads to neuropathological processes, including apoptotic neurons and gliosis.

All transgenic mice derived via pronuclear injections retain the ability to express mouse APP. It has been demonstrated that mouse amyloid peptides do not aggregate in solution nearly as well as the human amyloid peptides (Dyrks et al., *FEBS Lett.*, 324:231-36, 1993). It is likely that the mouse amyloid peptide interferes with the process of human amyloid aggregation. This may, in part, explain the necessity in the existing mouse AD model to greatly over-express human amyloid in a mouse brain to develop extra-cellular amyloid deposits.

The human APP gene locus encompasses a very large region (~400 Kb). Transgenic mice have been generated using YACs which appear to contain an intact human APP gene (Lamb et al., 1993; Pearson and Choi, *Proc. Natl. Acad. Sci.*, 30:10578-10582, 1993). But because gene regulatory elements have been identified at considerable distances from the proximal promoter of many genes (e.g., (Grosveld et al., 1987) and (Simonet et al., *J. Biol. Chem.*, 268:8221-8229, 1993)) there is no assurance that a given APP YAC clone contains all necessary APP gene regulatory elements. AD is a complex disease of aging, and the regulation of APP gene expression may play a critical role in the onset and progression of the disease. An accurate mouse model for AD may very well require the presence of critical APP gene regulatory elements which may be missing or altered in the YAC clones. In

addition, the YACs will integrate at random sites in the mouse chromosome after electroporation and expression of the human APP gene may be altered in a detrimental fashion due to "position" effects (see above).

YAC clones are inherently unstable and it can be very difficult to generate transgenic mouse lines where the gene locus resident on the YAC has remained intact. Furthermore, FAD mutations need to be introduced into the very large YACs via homologous recombination in yeast. Determining the stability and integrity of FAD-APP YACs require considerable effort (Lamb et al., 1993, Pearson and Choi, 1993).

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to a method of modifying a target nucleic acid. The target nucleic acid preferably comprises a genomic DNA sequence. The invention also relates to recombinant nucleic acid molecules which comprise a nucleotide sequence effective for homologous recombination at a predefined position of a gene and which is operably linked to a nucleotide coding sequence. Such recombinant nucleic acid molecules can be further combined with a vector sequence, a selectable marker, etc., to form a targeting vector useful for modifying a target nucleic acid, e.g., a genomic DNA sequence. The invention also relates to transgenic animals which comprise cells containing a recombinant gene, e.g., an APP gene or a presenilin gene, where the gene has been modified or engineered using the mentioned gene targeting vector. The transgenic animals are useful as animal models for diseases associated with the modified gene locus, e.g., Alzheimer's disease for the APP or presenilin genes.

An object of the invention is a novel gene targeting strategy that facilitates the introduction of one or more specific mutations into any gene in a single double reciprocal homologous recombination step, providing a clear advantage over other gene targeting approaches which use at least two

transfection and screening/selection steps. The gene targeting strategy preferably utilizes double reciprocal homologous recombination and a positive selectable marker gene to facilitate the insertion of gene segments or cDNA's (from the same or a heterologous host) into specific sites within the chromosome of a desired host cell, e.g., an embryonic stem (ES) cell derived from a rodent such as mouse. By the term "cDNA", it is meant a DNA which has been obtained by copying mRNA. The gene segments or cDNA's can be modified to encode one or more mutations. These gene-to-gene segments or gene-to-cDNA fusions, therefore, allow the introduction of one or more specific mutations into the coding sequence of the targeted gene. For some purposes, it may be preferable to employ a cDNA which is modified by the addition of other desired sequences, either coding or non-coding.

An aspect of the invention is a recombinant nucleic acid molecule comprising a nucleotide coding sequence, e.g., a cDNA, which is operably linked at its 5' or 3' terminus, or at both, to a nucleotide sequence which is effective to achieve homologous recombination. The invention also relates to a nucleotide sequence of a rodent APP gene such as a murine APP gene, or other mammal, which is effective to achieve homologous recombination at a predefined position in a target gene, operably linked to the 5' terminus, 3' terminus, or both, of a nucleotide sequence coding for at least one amino acid which is not naturally occurring at a specific amino acid position of the target gene. When the molecule comprises sequences at its 5' and 3' terminus which are homologous to the target gene, the molecule is effective to achieve homologous recombination with the target gene located, e.g., on a chromosome.

The term recombinant means a nucleic acid molecule which has been modified by the hand-of-man, e.g., comprising fragments of nucleic acid from different sources or a nucleic acid molecule from one source which has been engineered. Thus, the nucleic acid molecule is recombinant, e.g.,

because it comprises nucleotide sequences from a rodent (e.g., mouse) gene and a human gene or a synthetic (i.e., engineered) nucleotide sequence. Homologous recombination is a process in which nucleic acid molecules with similar genetic information line up side by side and exchange nucleotide strands. A nucleotide sequence of the recombinant nucleic acid which is effective to achieve homologous recombination at a predefined position of a target gene therefore indicates a nucleotide sequence which facilitates the exchange of nucleotide strands between the recombinant nucleic acid molecule at a defined position of a target gene, e.g., a mouse APP gene. The effective nucleotide sequence generally comprises a nucleotide sequence which is complementary to a desired target nucleic acid molecule (e.g., the gene locus to be modified), promoting nucleotide base pairing. Any nucleotide sequence can be employed as long as it facilitates homologous recombination at a specific and selected position along the target nucleic acid molecule. Generally, there is an exponential dependence of targeting efficiency on the extent or length of homology between the targeting vector and the target locus. Selection and use of sequences effective for homologous recombination is described, e.g., in Deng and Capecchi, *Mol. Cell. Bio.*, 12:3365-3371, 1992; Bollag et al., *Annu. Rev. Genet.*, 23:199-225, 1989; Waldman and Liskay, *Mol. Cell. Bio.*, 8:5350-5357, 1988.

The nucleotide sequence effective for homologous recombination can be operably linked to a nucleotide sequence, preferably comprising a nucleotide coding sequence, which is to be recombined into the desired target nucleic acid. For example, an aspect of the present invention is to replace all or part of the amino acids comprising exons 16, 17, and 18 of the APP gene with a cDNA coding for all or part of the corresponding amino acids of a human APP gene. This is achieved by attaching a part of the APP gene comprising a part of intron 15 and exon 16 to the 5' terminus of a human cDNA and a part of the APP gene comprising a part of intron 16 to the 3'

terminus of the cDNA to form a targeting vector. The APP gene segments are positioned with respect to the human cDNA in a way such that homologous recombination between them and the mouse gene will result in replacement of exons 16 through 18 with the cDNA. Such positioning, i.e., operable linkage, means that the mouse gene segment is joined to the cDNA whereby the homologous recombination function can be accomplished.

A nucleic acid comprising a nucleotide sequence coding without interruption means that the nucleotide sequence contains an amino acid coding sequence for a polypeptide, with no non-coding nucleotides interrupting or intervening in the coding sequence, e.g., absent intron(s) or the noncoding sequence, as in a cDNA.

An object of the present invention is to introduce modifications into genomic sequences, e.g., by introducing into or replacing a genomic sequence with a cDNA. Such cDNA can comprise one or more mutations, thereby facilitating the introduction of any desired nucleotide sequence into a target nucleic acid. The introduced nucleic acid, e.g., a DNA can particularly encode modifications in, or which span, two or more exons of a desired gene using only a single, double reciprocal homologous recombination event. In one embodiment, two independent point mutations can be introduced into a genomic sequence, where each point mutation is located in a different exon of the same gene. Thus, the coding sequence can be a nucleotide sequence which codes without interruption for an amino acid sequence, where the amino acid sequence is coded for by two or more exons in a naturally-occurring genomic (i.e., gene) sequence. This includes, e.g., a coding sequence for an amino acid sequence which is a cDNA, where the cDNA comprises amino acids coded for by separate exons of a naturally-occurring genomic sequence comprising exons and introns. By the phrase naturally-occurring genomic sequence, it is meant the gene structure as it occurs in nature. For example, a human APP gene contains 18 exons in a

naturally-occurring form which has been described. See, e.g., Yoshikai et al., *Gene*, 87:291-292, 1990. Other gene structures are also possible.

The introduction of point mutations via a replacement type vector has been described (Rubinstein et al., *Nucl. Acid Res.*, 21:2613-2617, 1993). Rubinstein et al. did not consider fusing genomic sequences with cDNA sequences to encode the gene product. Therefore, the gene targeting technology described by Rubinstein et al. would not succeed in introducing the Swedish-London and Swedish-714 stop double mutations into the mouse APP gene locus. The beta-amyloid domain resides on two separate exons (Lemaire et al., *Nucleic Acid Res.*, 17:517-522, 1989; Kang and Muller-Hill, *Biochem. Biophys. Res. Comm.*, 166:1192-2000, 1990). While the Swedish mutation and human amino acid differences reside on exon 16, the London mutation resides on exon 17. Lambda genomic clones are not large enough to encompass both exons (Lamb et al., *Nature Genetics*, 5:22-30, 1993). Therefore, the introduction of the double mutations into a host gene locus (e.g., a mouse APP gene) by the previously described gene-targeting approaches would require multiple gene targeting events utilizing two independent targeting vectors. Thus, it is recognized that in accordance with the present invention mutations which span sequences too large to fit into conventional vectors, targeting strategies, etc. (such as described in Lamb et al., 1993), e.g., two or more exons, can be introduced into genomic DNA by preparing targeting vectors comprising an intron effective for homologous recombination and a contiguous coding sequence, e.g., from the two or more exons.

The nucleotide coding sequence can code for at least one amino acid whose identity and/or position is not naturally-occurring in a target gene, e.g., a rodent (e.g., mouse) or non-human mammal gene. This means that the nucleotide coding sequence, when inserted into the target gene such that an open reading frame is formed with the target gene coding sequences,

contains at least one non-identical amino acid from the coding sequence of the unmodified target gene. This can mean amino acid substitution, deletion, or addition. In the examples below which illustrate, but in no way limit the invention, a nucleic acid coding for amino acids of a mouse APP gene are replaced by nucleic acid coding for amino acids of a human APP gene. At least 5 alternative splice forms of APP have been detected (reviewed in Beyreuther et al., Ann. NY Acad. Sci., *695*, 91-102 (1993)). The amino acids of a human APP gene means amino acid(s) identified as non-identical when the two APP gene sequences are compared. The amino acid numbering in the patent application refers to the largest alternative splice form of APP which consists of 770 amino acids. See, e.g., Kitaguchi et al., Nature *331*, 530-532 (1988); Tanaka et al., Biochem. Biophys. Res. Commun., *157*, 472-479 (1988). For example, the human amino acid sequence differs in the beta-amyloid domain are at positions 676, 681, and 684. The mouse APP gene contains a glycine at amino acid position 676, and a phenylalanine at amino acid position 681, and an arginine at amino acid position 684. A nucleotide coding sequence, which when inserted into an open reading frame of the mouse APP gene, comprising an arginine at amino acid position 676, a threonine at amino acid position 681, and/or a histidine at amino acid 684 is considered to contain three amino acid(s) whose identify is not naturally-occurring at an amino acid position (i.e., 676, 681, and/or 684) in the target mouse APP gene. See Figure 17 for other differences between the mouse and human APP polypeptide sequence.

A nucleic acid coding for at least one amino acid not naturally occurring in the targeted gene can also comprise, e.g., nucleotides which occur in a naturally-occurring human gene, such as naturally-occurring polymorphisms, alleles, or mutations which are discovered or identified in a natural population. By the term naturally-occurring, it is meant that the nucleic acid is obtained from a natural source, e.g., animal tissue and cells,

body fluids, tissue culture cells, forensic samples. Any other amino acid(s) can be incorporated, as well as, e.g., conservative and non-conservative amino acid substitutions, amino acid(s) obtained from other genes, non-naturally-occurring or engineered sequences, functional and/or selectable coding sequence domains.

In the examples, a mouse APP gene is targeted by the substitution of an amino acid found in a human APP gene. Numerous naturally-occurring mutations have been identified in non-murine APP genes. A nucleic acid according to the present invention can contain such mutations. Other modifications to the sequence can comprise mutations found in familial or genetic cases of disease, preferably Alzheimer's disease, Down's syndrome, or heredity cerebral hemorrhage with amyloidosis Dutch type (HCHWA-D). A nucleotide sequence coding for all or part of an amino acid sequence of a human APP gene can contain codons found in a naturally-occurring gene or transcript, or it can contain degenerate codons coding for the same amino acid sequences.

Preferred human APP amino acid sequences include: Swedish-FAD, KM(670,671)NL; London-FAD, V(717)I; Swedish/London-FAD, KM(670,671)NL, V(717)I; stop codon at position 714; Swedish-FAD, KM(670,671)NL, stop codon at position 714, etc. See Table 1.

An amino acid sequence of a human APP gene comprising a nucleotide sequence to be inserted into a targeted mouse APP gene preferably codes without interruption and comprises arginine at 676, threonine at position 681, histidine at position 684, or combinations thereof, in addition to other mutations and engineered codons.

The present invention also relates to nucleic acids which hybridize to a nucleic acid coding for an amino acid sequence of a human APP gene, preferably under stringent conditions. Such hybridizable sequences are

preferably not a naturally-occurring mouse APP nucleotide sequence; however, mutant mouse APP sequences can be included.

Hybridization conditions can be chosen to select nucleic acids which have a desired amount of nucleotide complementarity with the nucleotide sequence coding for all or part of an amino acid sequence of a human APP gene. A nucleic acid capable of hybridizing to such sequence, preferably, possesses 50%, more preferably, 70% complementarity, between the sequences. The present invention particularly relates to nucleotide sequences which hybridize to the nucleotide sequence coding for human APP amino acids under stringent conditions. As used here, "stringent conditions" means any conditions in which hybridization will occur where there is at least about 95%, preferably 97%, nucleotide complementarity between the nucleic acids. A nucleotide sequence hybridizing to the coding sequence will have a complementary nucleic acid strand, or act as a template for one in the presence of a polymerase (i.e., an appropriate nucleic acid synthesizing enzyme), which has a corresponding amount of nucleotide identity or similarity. The present invention includes both strands of nucleic acid, e.g., a sense strand and an anti-sense strand. Thus, it is understood that a nucleic acid comprising a nucleotide sequence hybridizing to the coding nucleotide sequence of amino acids of a human APP gene also represents a nucleic acid which possesses at least about 95%, preferably 97% nucleotide sequence identity.

According to the present invention, at least one amino acid not naturally-occurring in the targeted gene also includes amino acids selected from engineered or non-naturally-occurring sequences. In the examples, a mouse APP gene is modified by replacing mouse amino acids with amino acids which naturally occur in a human APP gene. However, the mouse APP gene can also be modified or engineered by the introduction of amino acids which are not based on a human APP gene, e.g., conservative or non-

conservative amino acids, cysteines, prolines, functional and/or selectable domains, etc.

Changes or modifications to the nucleotide coding sequence can be accomplished by any method available, including directed or random mutagenesis to a nucleic acid. These sequence modifications include, e.g., nucleotide substitution which does not affect the amino acid sequence (e.g., different codons for the same amino acid), replacing naturally-occurring amino acids with homologous or conservative amino acids, e.g. (based on the size of the side chain and degree of polarization), small nonpolar: cysteine, proline, alanine, threonine; small polar: serine, glycine, aspartate, asparagine; large polar: glutamate, glutamine, lysine, arginine; intermediate polarity: tyrosine, histidine, tryptophan; large nonpolar: phenylalanine, methionine, leucine, isoleucine, valine. In addition, it may be desired to change the codons in the sequence to optimize the sequence for expression in a desired host.

In addition to a gene segment effective for homologous recombination and coding sequence to be recombined, e.g., a recombinant nucleic acid molecule according to the present invention also can include selection markers, 3' regulatory sequences, regulatory sequences, restriction sites, vector sequences, and sequences and/or modification which enhance homologous recombination.

In order to identify cells which have integrated the nucleic acid molecule, it is desirable to include a selectable marker gene, e.g., neomycin resistance, gene HPRT gene, etc. A selectable marker gene codes for a product which can be directly or indirectly detected in a host in which it is expressed. Selectable marker genes and their use are widely used in molecular biology. When a neomycin resistance gene is utilized, cells having incorporated it can be selected by resistance to G418. A second selectable marker gene can also be incorporated into the vector, e.g., a herpes simplex

virus thymidine kinase gene. Any selectable genes routinely used in host cells can be used in the gene targeting vectors, including HSV TK, neo^r, hygromycin, histidinol, Zeocin (Invitrogen), HPRT, etc. Selectable genes can also be included to select against random integration events. Thus, selection for the first marker (e.g., by positive selection), and absence of the second marker (e.g., by negative selection), permits enrichment for transformed cells containing a modified target nucleic acid sequence, e.g., at the APP gene locus. The choice and arrangement of the selectable marker gene(s) in the recombinant nucleic acid molecule are as the skilled worker would know, e.g., described in U.S. Pat. No. 5,464,764 and Rubinstein et al., *Nucl. Acid Res.*, 21:2613-2617, 1993. A preferred recombinant nucleic acid comprises a selectable marker gene, e.g., a gene for neomycin resistance, in the mouse APP gene segment 3' to the cDNA. The selectable marker genes can be operably linked to regulatory sequences which control their expression, e.g., in a cell or tissue specific manner. Examples of such sequences are described, e.g., in U.S. Pat. No. 5,464,764.

In accordance with the present invention, 3' regulatory nucleotide sequences can be operably linked to a recombinant nucleic acid molecule. For example, it may be desirable to include a transcription termination signal and/or polyadenylation signal (e.g., AATAAA tandem repeat) at the 3' end of the nucleotide sequence to be inserted into the foreign gene. Generally, a selectable marker gene directly follows the transcription, termination and polyadenylations signals. Other sequences can also be included, e.g., nucleotide sequences which regulate the stability of a mRNA.

A recombinant nucleic acid can also comprise nucleotide sequences which affect expression of the gene into which it is combined, e.g., enhancers.

A recombinant nucleic acid molecule according to the present invention can also comprise all or part of a vector. A vector is a nucleic acid

molecule which can replicate autonomously in a host cell, e.g., containing an origin of replication. Vectors can be useful to perform manipulations, to propagate, and/or obtain large quantities of the recombinant molecule in a desired host. A skilled worker can select a vector depending on the purpose desired, e.g., to propagate the recombinant molecule in bacteria, yeast, insect, or mammalian cells. Examples of useful vectors include Bluescript KS+II (Stratagene). The following vectors are provided by way of example, Bacterial: pQE70, pQE60, pQE-9 (Qiagen), pbs., pD10, phagescript, psiX174, pbluescript SK, pbsks, pNH8A, pNH16a, pNH18Z, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene), pSVK3, pBPV, pMSG, pSVL (Pharmacia). However, any other vector, e.g., plasmids, viruses, or parts thereof, may be used as long as they are replicable and viable in the desired host. The vector can also comprise sequences which enable it to replicate in the host whose genome is to be modified. The use of such vector can expand the interaction period during which recombination can occur, increasing the targeting efficiency.

Recombinant nucleic acid molecules according to the present invention can also include sequences and modifications which decrease nonhomologous recombination events and/or enhance homologous recombination. For example, it has been found by Chang & Wislon, *Proc. Natl. Acad. Sci. USA*, 84:4959-63, 1987, that the addition of dideoxy nucleotides to the recessed termini of DNA molecules could enhance homologous recombination 6- to 7-fold relative to nonhomologous events.

Recombinant nucleic acid molecules according to the present invention can be prepared according to the various methods known to the skilled worker in the art, e.g., as mentioned in *Current Protocols in Molecular Biology*, Edited by F.M. Ausubel et al., John Wiley & Sons, Inc; and

Current Protocols in Human Genetics, Edited by Nicholas C. Dracopoli et al., John Wiley & Sons, Inc.

In accordance with the present invention, the novel gene-targeting can be used to modify any desired gene. Figure 15 illustrates several general strategies. Figure 15A shows a "typical" host gene with a DNA sequence consisting of a gene promoter, a series of exons (5 in this example). The exons are depicted as boxes. The gene can contain one or more exons. The line between the boxes (exons) represent the introns. The 5'-end of each intron contains a splice donor site which lies directly juxtaposed to the 3'-nucleotide to the preceding exon. The 3'-end of each intron contains a splice acceptor sequence which lies directly juxtaposed to the 5'-end of the neighboring exon. The 3'-end of the last exon contains a nonsense codon (designated as a stop) to terminate translation. This is followed by 3'-untranslated sequences which are present in the gene transcript and then a transcription termination and polyadenylation signal (designated polyA).

Figure 15B illustrates a targeted gene where a cDNA is inserted directly into an exon (exon 4 in this example) of the gene. Using an appropriately designed gene-targeting construct, any exon of a mouse gene can be targeted in this fashion. This is the approach used in the examples to generate the Swedish and/or London FAD-m/hAPP transgenic mice. The sequence of the cDNA is arranged so that the fusion between the gene and the cDNA creates an "in-frame" sequence that properly encodes the desired protein. The cDNA can be modified to encode one or more mutations (designated at *). The cDNA can be derived from transcripts from other genes of the same species or from genes from other species.

The cDNA is inserted into the mouse by homologous recombination. The recombination occurs between the targeted gene and an exogenously added gene-targeting construct or vector. The vector is preferably linearized. For homologous recombination to insert the cDNA into the proper location

and orientation within the targeted gene, the DNA components of the vector can be arranged in a specific manner. The cDNA is preferably positioned between nucleotide sequences which are homologous to specific locations of the targeted gene.

In the gene-targeting vector, there is preferably a gene segment comprising a nucleotide sequence corresponds substantially to an upstream (5'-flanking) region of the targeted gene. This segment comprises contiguous and sufficient upstream (5'-flanking) sequences of the targeted gene to allow efficient recombination to take place, i.e., a nucleotide sequence which is effective for homologous recombination. The segment can be followed by a portion of the targeted gene exon (exon 4 in this example).

In the gene-targeting vector, the sequence spanning the junction between the 3'-end of the targeted gene exon (exon 4 in this example) and the 5-end of the cDNA are arranged precisely in-frame to conserve the open reading frame to properly encode the desired gene product. In effect, the cDNA and the exon into which it is inserted become the terminal exon of the targeted gene. For proper termination and maturation of the transcript encoded by the targeted gene, transcription termination and polyadenylation signals (designated polyA) are positioned directly after the cDNA (and after the translation of stop codon). Directly following the transcription termination and polyadenylation signals, the gene targeting vector further comprises a selectable marker gene such as the neomycin resistance gene (designated neo^r) or the HPRT gene.

The gene targeting vector can further comprise a downstream (3'-flanking) region of homology to the targeted gene which is placed directly after the selectable marker gene, e.g., neo^r. The downstream region of homology can comprise contiguous gene sequences but can be any length of sequence providing it is sufficiently long to facilitate homologous recombination. The 5'-end of the downstream region of gene homology can

be located at any position proximal to the targeted gene as long as it lies downstream (3') of the mouse gene sequence which forms the junction between the targeted gene exon and the cDNA. After homologous recombination has taken place, the DNA sequence of the targeted gene positioned between the 3'-end of the upstream region of gene homology and the 5'-end of the downstream region of gene homology will have been deleted.

After homologous recombination takes place, exon sequences lying 5' of the exon/cDNA junction will encode the N-terminal portion of the gene product while the cDNA sequences lying 3' of the exon/cDNA junction will encode the C-terminal portion of the gene product.

Figure 15C illustrates a targeted gene where a cDNA is inserted directly into an intron (intron 3 in this example) of the targeted gene. Using an appropriately gene-targeting construct, any intron of a gene could be targeted in this fashion.

The sequence of the cDNA is arranged so that it functions as the terminal exon of the targeted gene. To form an open reading frame between the targeted and human coding sequence, the codon reading-frame of the cDNA sequence is positioned in-frame with the codon reading-frame of the nearest upstream (5') exon (exon 3 in this example). For proper splicing of messenger RNA to occur, a functional splice acceptor site immediately preceding (5') the cDNA can be included. Thus, after splicing of the exon with the cDNA, a resultant transcript from the targeted gene will encode the desired gene product. As mentioned above, a cDNA from various sources can be utilized and it can be modified to encode mutations. The arrangement of the gene targeting vector is as described above.

Figure 15D illustrates a targeted gene where a gene segment from another the same or different species (designated as foreign gene segment) is inserted directly into an intron (intron 3 in this example) of the targeted gene.

Using an appropriately gene-targeting construct, any intron of a gene can be targeted in this fashion. In this example, the sequence of the foreign gene segment contains normal exons and introns from another gene. The sequences of the gene-targeting construct are arranged such that the foreign gene segment functions as the terminal set of exons for the targeted gene. The codon reading-frame of the exons of the foreign gene segment can be arranged in-frame with the codon reading-frame of the nearest upstream (5') exon (exon 3 in this example) to form a complete open-reading frame. For proper splicing of messenger RNA to occur, a functional splice acceptor site immediately preceding the 5' exon of the foreign gene segment can be included. Thus, after splicing of the exon with the exons of the foreign gene segment, the transcript from the targeted gene will encode the desired gene product. The foreign gene segment can be obtained from various sources, as desired, and can be engineered to encode one or more mutations. The arrangement of a gene targeting vector is described above.

Another aspect of the present invention relates to host cells comprising a recombinant nucleic acid of the invention. A cell into which a nucleic acid is introduced is a transformed cell. Host cells include, mammalian cells, e.g., rodent, murine Ltk-, murine embryonic stem cells, COS-7, CHO, HeLa, insect cells, such as Sf9 and Drosophila, bacteria, such as E. coli, Streptococcus, bacillus, yeast, fungal cells, plants, embryonic stem cells (e.g., mammalian, such as mouse or human), neuronal cells (primary or immortalized), e.g., NT-2, NT-2N, PC-12, SY-5Y, neuroblastoma. See, also *Methods in Enzymology*, Volume 185, ed., D.V. Goeddel. A nucleic acid can be introduced into the cell by any effective method including, e.g., calcium phosphate precipitation, electroporation, injection, DEAE-Dextran mediated transfection, fusion with liposomes, and viral transfection. When the recombinant nucleic acid is present in a host cell, it is preferably

integrated by homologous recombination into a chromosome residing in the host cell.

The present invention also relates to a recombinant nucleic acid coding for a recombinant polypeptide, which nucleic acid is a product of the gene which has been modified by the gene targeting vector. A gene can code for different nucleic acid transcripts, depending on splicing, where it is expressed, etc. All such nucleic acids are a product of the recombinant gene and thus relate to the present invention. Such nucleic acids can code for recombinant polypeptides which are also an object of the present invention. The recombinant polypeptides can be used, e.g., as antigens to generate specific antibodies as diagnostic, research, and therapeutic tools.

A recombinant nucleic acid and a recombinant polypeptide can incorporate at least one amino acid or coding sequence thereof from a heterologous species. If, e.g., a non-human mammal sequence contains at least one amino acid of a human sequence, the modified sequence is described as "humanized." By "humanized" it is meant, e.g., a mouse polypeptide containing one or more amino acids which are present in the human polypeptide (and which differ from the amino acids present in the mouse gene).

Thus, in the examples, humanized mouse APP nucleic acids and polypeptides were created by substituting a human amino acid for a mouse amino acid at corresponding locations. A recombinant nucleic acid can be an unprocessed RNA transcript comprising introns or it can comprise a nucleotide sequence coding without interruption for amino acids, e.g., where the nucleic acid is a modified APP gene, it can code for amino acids 1-770, 1-713, 1-751, and 1-695. For example, a nucleic acid coding for a recombinant APP polypeptide can be a transcript from an APP gene modified in accordance with the present invention, e.g., by homologous recombination with a human cDNA and a mouse gene. The recombinant nucleic acid can

comprise mutations in the APP gene, e.g., Swedish-FAD, London-FAD, etc., as described above.

The present invention also relates to a non-human transgenic animal, preferably a mammal, more preferably a rodent such as a mouse, which comprises a gene, which has been engineered employing a recombinant nucleic acid according to the present invention. Generally, a transformed host cell, preferably a totipotent cell, whose endogenous gene has been modified using a recombinant nucleic acid as described above is employed as a starting material for a transgenic embryo. The preferred methodology for constructing such a transgenic embryo involves transformed embryonic stem (ES) cells employing a targeting vector comprising a recombinant nucleic acid according to the invention. A particular gene locus, e.g., APP, is modified by targeted homologous recombination in cultured ES cells employing a targeting vector comprising a recombinant nucleic acid according to the invention. The ES cells are cultured under conditions effective for homologous recombination. Effective conditions include any culture conditions which are suitable for achieving homologous recombination with the host cell chromosome, including effective temperatures, pH, medias, additives to the media in which the host cell is cultured (e.g., for selection, such as G418 and/or FIAU), cell densities, amounts of DNA, culture dishes, etc. Cells having integrated the targeting vector are selected by the appropriate marker gene present in the vector. After homologous recombination has been accomplished, the cells contain a chromosome having a recombinant gene. In a preferred embodiment, this recombinant gene contains host genomic sequences (e.g., mouse) fused to a donor cDNA (e.g., human). The cDNA can contain multiple mutations, etc., which are not naturally-occurring in the target gene. No further gene engineering steps are necessary. Thus, in accordance with the present invention, a single step has resulted in a modified gene containing as many modified sequences as

desired. Another aspect of the present invention involves employing a cDNA with sufficient nucleotide sequence dissimilarity between it and the native target gene sequence to avoid inappropriate intra-recombination and inter-recombination events, subsequent to the first gene targeting step.

The transformed or genetically modified cells can be used to generate transgenic non-human mammals, e.g., rodents (such as mice or rats), by injection into blastocysts and allowing the chimeric blastocysts to mature, following transfer into a pseudopregnant mother. See, e.g., *Teratomacarcinoma and Embryonic Stem Cells: A Practical Approach*, E.J. Robertson, ed., IRL Press. Various stem cells can be used, as known in the art, e.g., AB-1, HM-1 D3, CC1.2, E-14T62a, preferably ES cell line G1 derived from inbred mouse strain 129/SvEvT.

In accordance with the present invention, a transformed cell contains a recombinant gene integrated into its chromosome at the targeted gene locus. A targeting vector which comprises sequences effective for homologous recombination at a particular gene locus, when introduced into a cell under appropriate conditions, will recombine with the homologous sequences at the gene locus, introducing a desired gene segment (e.g., a cDNA) into it. When recombination occurs such that insertion results, the nucleic acid is integrated into the gene locus. The gene locus can be the chromosomal locus which is characteristic of the species, or it can be a different locus, e.g., translocated to a different chromosomal position, on a supernumerary chromosome, on an engineered "chromosome," etc. In the examples described below, the sequences of the human APP gene are integrated by homologous recombination into the normal APP gene loci on murine chromosome 16. By recombinant, it is meant that the nucleotide sequences come from different sources, e.g., mouse and human.

A transgenic non-human mammal comprising a recombinant gene, which when mutant results in Alzheimer's disease, can express the gene in an

amount effective to produce neuronal cell degeneration and/or apoptosis. The gene can also be expressed in an amount effective to cause a behavioral or cognitive dysfunction, wherein the dysfunction is conferred by the recombinant gene. Such gene can be, e.g., PS1, PS2, S182 (e.g., Sherrington et al., *Nature*, 375:754-760, 1995), STM2, E5-1, apoliprotein E, apoptosis genes such as ALG-1 to -6 (Vito et al., *Science*, 271:521, 1995), Bcl-2/Bax gene family, etc.

A transgenic non-human animal according to the present invention can comprise one or more genes which have been modified by genetic engineering. For example, a transgenic animal comprising an APP gene which has been modified by targeted homologous recombination in accordance with the present invention can comprise other mutations, including modifications at other gene loci and/or transgenes, including PS1, PS2, S182 (e.g., Sherrington et al., *Nature*, 375:754-760, 1995), STM2, E5-1, apoliprotein E, apoptosis genes such as ALG-1 to -6 (Vito et al., *Science*, 271: 521, 1995), Bcl-2/Bax gene family, etc. Modifications to these gene loci and/or introduction of transgenes can be accomplished in accordance with the methods of the present invention, or other methods as the skilled worker would know, e.g., by pronuclear injection of recombinant genes into pronuclei of one-cell embryos, incorporating an artificial yeast chromosome into embryonic stem cells, gene targeting methods, embryonic stem cell methodology. See, e.g., U.S. Patent Nos. 4,736,866; 4,873,191; 4,873,316; 5,082,779; 5,304,489; 5,174,986; 5,175,384; 5,175,385; 5,221,778; Gordon et al., *Proc. Natl. Acad. Sci.*, 77:7380-7384 (1980); Palmiter et al., *Cell*, 41:343-345 (1985); Palmiter et al., *Ann. Rev. Genet.*, 20:465-499 (1986); Askew et al., *Mol. Cell. Bio.*, 13:4115-4124 (1993); Games et al. *Nature*, 373:523-527 (1995); Valancius and Smithies, *Mol. Cell. Bio.*, 11:1402-1408 (1991); Stacey et al., *Mol. Cell. Bio.*, 14:1009-1016 (1994); Hasty et al.,

Nature, 350:243-246 (1995); Rubinstein et al., *Nucl. Acid Res.*, 21:2613-2617 (1993).

A recombinant nucleic acid molecule according to the present invention can be introduced into any non-human mammal, including a rodent, mouse (Hogan et al., *Manipulating the Mouse Embryo: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1986), rat, pig (Hammer et al., *Nature*, 315:343-345, 1985), sheep (Hammer et al., *Nature*, 315:343-345, 1985), cattle or primate. See also, e.g., Church, *Trends in Biotech.* 5:13-19, 1987; Clark et al., *Trends in Biotech.* 5:20-24, 1987; and DePamphilis et al., *BioTechniques*, 6:662-680, 1988.

A transgenic non-human animal and a recombinant nucleic acid molecule according to the present invention is useful as described in U.S. Pat. Nos. 5,304,489, 5,221,778, 5,175,385, 5,175,384, 5,175,383, 5,087,571, 5,082,779, 4,736,866, 4,873,191, and other transgenic animal patents. For example, a recombinant nucleic acid molecule comprising a coding sequence for at least one amino acid of a human APP gene is useful as a hybridization probe for detecting and diagnosing Alzheimer's disease, e.g., nucleotide variations and genetic polymorphisms present in a nucleic acid can be detected in accordance with various methods, e.g., U.S. Pat. 5,468,613; Conner et al., *Proc. Natl. Acad. Sci.* 80, 78 (1983); Angelini et al., *Proc. Natl. Acad.*, 83, 4489 (1986); Myers et al., *Science*, 230, 1242 (1985). The nucleic acid can also be operably linked to an expression control sequence to produce polypeptide encoded by it. The operable linkage of nucleic acid and expression control sequence can be introduced into a desired host, and cultured under conditions effective to achieve expression of a polypeptide coded for the nucleic acid. An expression control sequence is similarly selected for host compatibility and a desired purpose, e.g., high copy number, high amounts, induction, amplification, controlled expression. Other sequences which can be employed, include, enhancers such as from

SV40, CMV, inducible promoters, neuronal specific elements, or sequences which allow selective or specific cell expression, such as in neuronal cells, glial cells, etc. The expression control sequence includes mRNA-related elements and protein-related elements. Such elements include promoters, enhancers (viral or cellular), ribosome binding sequences, transcriptional terminators, etc. An expression control sequence is operably linked to a nucleotide coding sequence when the expression control sequence is positioned in such a manner to effect or achieve expression of the coding sequence. For example, when a promoter is operably linked 5' to a coding sequence, expression of the coding sequence is driven by the promoter. The resulting polypeptides can be used to generate antibodies for diagnostic purposes, etc. The operable linkage with an expression control sequence can also occur *in situ* as a result of homologous recombination at the desired gene locus, e.g., a mouse APP gene.

A further aspect is the expression of a modified mRNA and polypeptides encoded by a recombinant nucleic acid molecule of the present invention in a transgenic animal, preferably a non-human mammal, as a model for diseases associated with the gene, e.g., the APP, PS1, and PS2 genes with Alzheimer's disease (AD), Down's syndrome, and heredity cerebral hemorrhage with amyloidosis Dutch type (HCHWA-D). Expression of a modified gene product in a transgenic non-human mammal and its consequent phenotype can therefore be used as a model for diseases and pathologies, e.g., as an AD model for genes associated with Alzheimer's disease. As described in the examples below, a mouse APP gene is modified by the introduction of mutations which are associated with an Alzheimer's phenotype in humans. Transgenic mice comprising cells which contain such a modified or recombinant APP gene can be used to design therapies. For example, active agents, e.g., synthetic, organic, inorganic, or nucleic acids based molecules, can be administered to a transgenic non-human mammal

according to the present invention to identify agents which either inhibit, prevent, and/or reduce the appearance of an Aβ peptide in the brain, the AD pathology, neurodegeneration, apoptosis, cognitive deficits, and/or behavioral symptoms, etc. Thus, another aspect of the invention is to provide a method to assist in the advancement of the treatment and/or prevention of the aforementioned symptoms (e.g., neurodegeneration or apoptosis) caused by the APP gene, or a fragment thereof. Other genes and therapies can be used analogously.

Such a mammal model can also be used to assay for agents, e.g., zinc, and factors, e.g., environmental, which exacerbate and/or accelerate the diseases. See, e.g., Bush et al., *Science*, 265:1464-1467, 1994. A transgenic non-human animal can also be useful as pets, food sources (e.g., mice for snakes), in toxicity studies, etc.

Moreover, a non-human mammal containing a recombinant nucleic acid according to the present invention can be used in a method of screening a compound for its effect on a phenotype of a mammal, preferably a mouse, where the phenotype is conferred by the recombinant nucleic acid. By "phenotype," it is meant, e.g., a collection of morphological, physiological, biochemical, and behavioral traits possessed by a cell or organism that results from the interaction of the genotype and the environment. A phenotype can be behavioral, e.g., occurrence of seizures or cognitive performance, or it can be physiological and/or pathological, e.g., occurrence of neuronal cell degeneration, neuronal cell apoptosis, accumulation of Aβ peptide in the brain of the mammal, altered carboxy-terminal processing of the APP polypeptide, etc. According to such a method of detection, a compound can be administered to a mammal containing a modified APP gene and then the existence of an effect on the phenotype of the mammal can be determined. Observation can be accomplished by any means, depending on the specific phenotype which is being examined. For example, the ability of a test

compound to suppress a behavioral phenotype can be detected by measuring the latter phenotype before and after administration of the test compound.

The invention also relates to a transgenic non-human mammal comprising cells that contain a recombinant gene modified by a gene targeting vector. For example, such recombinant gene or nucleic acid can code for a humanized mouse polypeptide comprising at least one amino acid coded for by a human gene, e.g., where the gene is the APP, PS1, or PS2 gene. In the case of the APP gene, the gene can code for, e.g., amino acids 1-665 of a mouse APP gene and amino acids 666-770 of a human APP gene, and having a phenotype conferred by the modified gene, e.g., accumulation of Aβ peptide or other related peptide in the brain, abnormal processing of the APP polypeptide, etc.

The level of expression of the recombinant gene can be any amount which can produce a phenotype in the non-human mammal, which phenotype can be distinguished from mammals which do not possess the modified gene locus, i.e., a control mammal, e.g., an amount effective to produce neuronal cell degeneration and/or apoptosis and/or an amount effective to cause a behavioral and/or cognitive effect or dysfunction where the gene is an alzheimer's disease associate gene.

A non-human mammal containing a modified APP gene can also be characterized by accumulation of the Aβ peptide in its brain. The accumulation can be in any quantity which is greater than that observed in mammals not containing the modified gene locus. The phenotype conferred by the modified APP gene can occur before or after accumulation can be detected. The expression and/or accumulation of the APP polypeptide, and its processed derivatives, and the nucleic acids which encode it, can be measured conventionally, e.g., by immunoassay or nucleic acid hybridization, either *in situ* or from nucleic acid isolated from host tissues.

The identification of agents which prevent and/or treat symptoms associated with expression of the modified gene can be determined routinely. For example, an active agent can be administered to a transgenic mammal comprising a modified gene according to the present invention and then its effect on a behavior or pathology, e.g., Aβ deposition in the brain, apoptosis, and/or neurodegeneration, can be determined. The agent can be administered acutely (e.g., once or twice) or chronically by any desired route, e.g., subcutaneously, intravenously, transdermally, or intracathically. The formulation of the agent is conventional, see, e.g., *Remington's Pharmaceutical Sciences*, Eighteenth Edition, Mack Publishing Company, 1990. In a test, e.g., an agent can be administered in different doses to separate groups of transgenic mammals to establish a dose-response curve to select an effective amount of the active agent. Such effective amount can be extrapolated to other mammals, including humans.

The transgenic mammal, preferably a mouse, according to the present invention therefore permits the testing of a wide variety of agents and therapies. In AD, for example, a number of different agents have been identified which affect the cognitive dysfunction associated with the diseases, e.g., cholinergic agents, such as muscarine agonists, acetylcholinesterase inhibitors, acetylcholine precursors, biogenic amines, nootropics, angiotensin converting enzyme (ACE), and vitamin E. In addition, agents which regulate APP or Aβ expression, Aβ deposition, and physiological changes associated with Aβ expression and deposition can also be identified, e.g., calcium homeostasis, inflammation, neurofibrillary tangles. See, e.g., Pavia et al., *Annual Reports of Medicinal Chemistry*, 25:2129, 1989; John et al., *Annual Reports of Medicinal Chemistry*, 28:197-203, 1993. Additionally, active agents which block apoptosis, e.g., free radical scavengers, such as glutathionines, can be administered. Such effects on AD can be assayed in either behavioral or physiological and/or histological studies.

For example, spatial learning and memory abilities in mice can be tested in a Morris water maze. See, e.g., Yamaguchi et al., *NeuroReport*, Vol. 2, 781-784 (1991). Additionally, other behavioral tests can be used, e.g., Swim Test, Morris et al., Learning and Motivation, 12, 239-260, 1981; Open-field test, Knardahl et al., *Behav. Neurol. Biol.* 27, 187-200, 1979; and tests and models used routinely, e.g., in mice, rats, and other rodents.

According to the present invention, differences in, e.g., levels of expression, cellular localization, and/or onset of expression of the recombinant gene can be used to model a disease, e.g., AD and other diseases associated with APP expression and the differing stages and progressions of the disease, e.g., cell degeneration, cell death, astrogliosis, and/or amyloidosis. Having a range of expression phenotypes can be useful to identify different therapies and drug treatments and also diagnostically to identify a disease's progression. For example, the specific treatments can depend on the region of the brain in which an APP peptide is expressed, how much of it is expressed, and its temporal progression of expression. Thus, mammals having different phenotypes can be used as models for determining therapies which are selective for different stages of the disease and for studying disease progression and intervention.

DESCRIPTION OF THE FIGURES

Figure 1. Schematic of p35A; mouse APP exon 16 genomic clone

The ~15 Kb Not I genomic fragment (shown) was isolated from the lambda clone 35A and cloned into the Not I site of Bluescript II SK+. Exon 16 is indicated and begins approximately 9.5 Kb from the 5'-end of the genomic fragment. The indicated restriction enzyme recognition sites were placed for reference.

Figure 2. Restriction Map pMTI-2396

pMTI-2396 contains mouse APP exon 16 and was derived from the ~5.5 Kb NcoI fragment from p35A (NcoI at position 7645 to NcoI at position 13176, Figure 1). The 5.5 Kb NcoI fragment was inserted into NcoI-modified Bluescript II SK+ at the NcoI site. All recognition sites for the indicated restriction enzymes are designated. Sequence from positions 29 to 5560 were derived from the mouse APP gene and the remaining sequences were derived from Bluescript II SK+.

Figure 3. Restriction map of pRA3

pRA3 contains mouse APP intron 15 sequences and was derived from the ~3 Kb NcoI fragment from p35A (NcoI at position 4816 to NcoI at position 7645, Figure 1). The 3 Kb fragment was inserted into NcoI-modified Bluescript II SK+ at the NcoI site. All recognition sites for the indicated restriction enzymes are designated. Sequence from positions 29 to 2858 were derived from the mouse APP gene and the remaining sequences were derived from Bluescript II SK+.

Figure 4. Restriction map of pN2C4

pN2C4 contains mouse APP intron 16 sequences and was derived from the ~1.9 Kb NcoI fragment from p35A (NcoI at position 13176 to NcoI at position 14992, Figure 1). The ~1.9 Kb fragment was inserted into NcoI-modified Bluescript II SK+ at the NcoI site. All recognition sites for the indicated restriction enzymes are designated. Sequence from positions 29 to 1845 were derived from the mouse APP gene and the remaining sequences were derived from Bluescript II SK+.

Figure 5. Restriction map of pMTI-2398; Swedish-FAD targeting vector

The mouse APP intron 15 and exon 16 sequences encompass positions 30 to 1960 (BglII site). The human APP cDNA and genomic polyadenylation sequences are contained in sequences between positions 1960 and ~4556. The neomycin resistance gene lies between positions ~4556 and ~6460. Mouse APP intron 16 sequences are contained between positions ~6460 and 9872. The Bluescript II SK+ sequences are between positions ~9872 and ~30. All recognition sites for the indicated restriction enzymes are designated.

Figure 6. Restriction map of pMTI-2453; London-FAD targeting vector

The HSV TK gene is located between positions ~17 and ~2893. The mouse APP intron 15 and exon 16 sequences encompass positions ~2906 to 4835 (BglII site). The human APP cDNA and genomic polyadenylation sequences are contained in sequences between positions 4835 and ~7452. The neomycin resistance gene lies between positions ~7452 and ~9323. Mouse APP intron 16 sequences are contained between positions ~9323 and 12750. The Bluescript SK+II sequences are between positions ~12750 and ~37. All recognition sites for the indicated restriction enzymes are designated.

Figure 7. Restriction map of pMTI-2454; Swedish/London-FAD targeting vector

The HSV TK gene is located between positions ~17 and ~2893. The mouse APP intron 15 and exon 16 sequences encompass positions ~2906 to 4835 (BglII site). The human APP cDNA and genomic polyadenylation sequences are contained in sequences between positions 4835 and ~7452. The neomycin resistance gene lies between positions ~7452 and ~9323. Mouse APP intron 16 sequences are contained between positions ~9323 and 12750. The Bluescript II SK+sequences are between positions ~12750 and ~37. All recognition sites for the indicated restriction enzymes are designated.

Figure 8. Restriction map of pMTI-2455 (Swedish-FAD APP713 targeting vector)

The HSV TK gene is located between positions ~17 and ~2893. The mouse APP intron 15 and exon 16 sequences encompass positions ~2906 to 4835 (BglII site). The human APP cDNA and genomic polyadenylation sequences are contained in sequences between positions 4835 and ~7452. The neomycin resistance gene lies between positions ~7452 and ~9323. Mouse APP intron 16 sequences are contained between positions ~9323 and 12750. The Bluescript II SK+ sequences are between positions ~12750 and ~37. All recognition sites for the indicated restriction enzymes are designated.

Figure 9. Oligonucleotides

Oligonucleotides are designated in the 5' to 3' direction.

Figure 10. Schematic outline of m/hAPP gene products produced in transgenic mouse lines.

The protein m/hAPP exhibits amino acid sequence identity with mouse APP with the exception of those residues indicated by (asterisks, see text above). m/hAPP protein spans the membrane once as indicated. The bA4 peptide region (indicated by red) partially resides in the transmembrane and extracellular domains. The APP751 alternative splice form of APP has the 56 amino acid Kunitz protease inhibitor domain while the APP770 splice form of the protein has both the Kunitz and the 19 amino acid OX domains. The APP695 alternative splice form of APP contains neither Kunitz nor OX domains. Other splice forms are not indicated. There are two possible N-linked glycosylation sites (CHO) in the extracellular domain of APP. A highly negatively-charged domain and a cysteine-rich domain are symbolized by a minus sign and S-S bridges respectively. The signal peptide (SP) is located at the N-terminus (see Unterbeck et al.).

Figure 11. Gene-targeting strategy: Construction of targeting vectors.

The schematic of the Nco I APP gene fragment represents the ~5.5 Kb NcoI mouse APP gene fragment in pMTI-2396 (Figure 2). The regions indicated in red represent the coding sequences for mouse b-amyloid domain. The schematic for the targeting vector represents the linearized (using AscI) DNA from clone pMTI-2454 (Figure 7). The targeting vectors for pMTI-2453 (Figure 6) and pMTI-2455 (Figure 8) are identical to pMTI-2454 with the exception of the FAD mutation and the orientation of the HSV TK gene (see text). pMTI-2398 is similar to pMTI-2454 with the exception of FAD mutation and the absence of the HSV TK gene (see text). The FAD mutations are indicated by black asterisks and the mutations to "humanize" the b-amyloid domain are indicated by green asterisks. The neomycin resistance gene is designated by neo^r and Bluescript II SK+ sequences are designated by BSSK+.

Figure 12. Gene-targeting strategy: Homologous recombination.

The linearized targeting vector (Figure 11) was electroporated into ES cells. Homologous recombination occurred between mouse APP sequences contained in the targeting vector and mouse APP genomic sequences on chromosome 16. The resulting targeted m/hAPP gene locus is schematically shown. The FAD mutations are indicated by asterisks and the mutations to "humanize" the b-amyloid domain are indicated by asterisks.

Figure 13. Gene-targeting strategy: Targeted m/hAPP gene locus.

The comparison of the mouse APP and targeted m/hAPP gene loci is shown schematically. In normal mouse, the b-amyloid, transmembrane, and cytoplasmic domains of APP are encoded by mouse APP exons 16, 17, and 18. In the case of the targeted m/hAPP gene locus, the b-amyloid, transmembrane, and cytoplasmic domains of m/hAPP are encoded by human cDNA sequences. The remainder of m/hAPP is encoded by mouse APP exons 1 through 15. The FAD mutations are indicated by asterisks and the mutations to "humanize" the b-amyloid domain are indicated by asterisks.

Figure 14. Strategy for Southern-blot detection of ES cells having a targeted m/hAPP gene locus containing the Swedish-FAD mutation (e.g.; transgenic lines ES5007 and ES5103).

The schematics for the mouse and m/hAPP loci are indicated. The restriction enzymes XbaI and NcoI are designated by X and N respectively. The box represents human APP cDNA and genomic sequences while the box represents the neomycin resistance gene.

Figure 15. Gene Targeting Strategies

A. Normal Gene

B. Targeted gene. Fusion of a gene with cDNA (in-frame fusion of mouse exon sequences with cDNA). * represents one or more mutations.

C. Targeted gene. Fusion of a target gene with cDNA (cDNA is inserted into a mouse intron (intron 3 for example). The cDNA is directly preceded by a splice acceptor site. The sequence of the insert is formatted so that splicing of the 3'-sequence of the exon (exon 3 for example) with the 5'-sequence of the cDNA will create a mature transcript encoding the appropriate gene product). * represents one or more mutations.

D. Targeted gene. Fusion of a targeted gene with a foreign (same or different species) gene segment including one or more exons inserted into the intron of the targeted gene. The sequence of the insert is formatted so that splicing of the 3'-sequence of the mouse exon (exon 3 for example) with the 5'-sequence of the other mouse gene or species exon (exon 4' for example) will create a mature transcript encoding the appropriate gene product). * represents one or more mutations.

Figure 16. Amino acid sequence of human APP.

Figure 17. Sequence of mouse exon 16 locus

Figure 18. Sequence of pMTI-2398 (Swedish-FAD APP targeting vector)

Figure 19. Sequence of pMTI-2453 (London-FAD APP targeting vector)

Figure 20. Sequence of pMTI-2454 (Swedish/London-FAD APP targeting vector)

Figure 21. Sequence of pMTI-2455 (Swedish-FAD APP713 targeting vector)

Figure 22. Sequence of APP genomic clone containing human APP polyadenylation signals.

EXAMPLES

Four independent lines of transgenic mice (lines ES5007, ES5103, ES5401 and ES5403) have been created via a novel gene targeting technique applied to embryonic stem cells. In each line, the mouse APP gene has been modified to encode a mouse/human hybrid APP (m/hAPP) where amino acid residues 666-770 of APP770 are now encoded by human cDNA sequences instead of mouse genomic exons (exons 16, 17, and 18). Within these residues only three amino acid differences exist between the mouse and human proteins (Gly(676) to Arg, Phe(681) to Thr, and Arg(684) to His). This exon-cDNA fusion gene, therefore, encodes an APP containing a "humanized" beta-amyloid domain (aa residues 672 to 712).

In each transgenic mouse line, the human cDNA sequences have been modified to introduce one or more mutations proximal to the "humanized" beta-amyloid domain. In transgenic mouse line ES5007, m/hAPP has been mutated to include the "Swedish"-FAD mutation (KM to NL, positions 670 and 671)(Cai et al., 1993, Citron et al., 1994). Transgenic mouse lines ES5401 and ES5403 encode m/hAPP which have been mutated to include the "London"-FAD mutation (V to I, position 717) (Suzuki et al., 1994, Gravina, 1995). Transgenic mouse line ES5103 encodes m/hAPP which has been mutated to include both "London" and "Swedish" FAD mutations. A fifth transgenic mouse line ES5215 can be produced which encodes m/hAPP that has been mutated to include both the "Swedish" FAD mutation and a premature stop codon (T to stop at position 714). With the exception of the changes mentioned above, the remainder of the m/hAPP sequences are identical to those found in normal mouse APP.

We have shown that the targeted Swedish-FAD m/hAPP and Swedish/London-FAD m/hAPP genes express m/hAPP protein at levels approaching those observed for mouse APP in brain.

Notably, we have observed that the Swedish FAD mutation alters significantly the proteolytic processing of APP resulting in differences in the appearance of C-terminal fragments. The observed changes in processing is consistent with the Swedish-FAD mutation inducing the beta-secretase cleavage site to be utilized predominately over the alpha-secretase cleavage site as previously observed in cell culture experiments (see below).

Messenger RNA from the Swedish-FAD m/hAPP gene was found be abundantly expressed in the brain from homozygous ES5007 mice as well. The amount of Swedish-FAD m/hAPP mRNA in homozygous ES5007 brain was determined to be approximately 55% of the mAPP mRNA levels observed in control mouse brain. In concordance, the APP mRNA levels in heterozygous ES5007 mouse brain were found to be approximately 75% of the level observed in control mouse brain.

The reverse transcriptase-PCR (rtPCR) technique was used to identify mouse APP and Swedish-FAD m/hAPP transcripts in mouse brain. Homozygous ES5007 mice were found to express mRNA exclusively from the targeted Swedish-FAD m/hAPP gene. No mRNA species containing sequences from mouse APP exons 16, 17, or 18 was detected in homozygotes. As would be expected, heterozygous ES5007 mice were found to express mRNA transcripts from both normal mouse and Swedish-FAD APP alleles.

Western-blot analyses have demonstrated that Swedish-FAD m/hAPP and Swedish/London-FAD protein is expressed in the brain of ES5007 and ES5130 mice, respectively. Swedish-FAD m/hAPP protein is expressed in the brain of homozygous ES5007 mice at approximately 87% of the level observed for mouse APP in non-transgenic mice (n = 4).

Retrieving mouse APP exon 16 from genomic library

Phage lifts: The mouse 129 genomic library from Stratagene (cat#946308) was titered and plated out 20 150 mm LB plates containing ~50,000 phage/plate. Duplicate lifts were made from each plate using Amersham Hybond-N+ nylon membranes. The plates were refrigerated for several hours to ensure the top agar was hardened. The membranes were placed atop the plaques and left on for 5 minutes. The membranes were lifted off the plates and placed plaque-side up on 3MM paper saturated with denaturation solution (0.1 M NaOH, 1.5 M NaCl) for 5 minutes. The membranes were transferred briefly to dry 3MM paper to absorb the excess solution and then placed on 3MM paper saturated with neutralizing solution (0.2 M Tris-Cl pH 7.5, 2X SSC) for 5 minutes. The membranes were rinsed by placing them on 3MM paper saturated with 2X SSC for 5 minutes and then air dried. A digoxigenin-labeled mouse specific APP exon 16 probe of 93 bp was generated using PCR (from nt 1877 to 1969 in sequence MUSABPPA, accession #M18373).

PCR assay: In a 50 μl total reaction volume was added 1 μg genomic mouse tail DNA, 5 μl 10X PCR buffer (Perkin Elmer cat#N808-0006), 5 μl 2 mM dATP, dCTP, dGTP mix, 5 μl 1.3 mM dTTP, 3.5 μl 1 mM digoxigenin-11-dUTP, 3 μl 100 ng/ml oligonucleotide mix of KC65 (5'GTTCTGGGCTGACAAACATC3') and KC66 (5'GATGGCGGACTTCAAATCCTG3'), and 2.5 units AmpliTaq (Perkin Elmer cat#N808-0070). The reaction was run in a Perkin Elmer turbo 9600 thermal cycler. The parameters of the run were as follows: one cycle at 94°C for one minute, 36 cycles at 94°C for 30 seconds-56°C for 50 seconds-70°C for two minutes, maintain at 10°C indefinitely. Four individual PCR reactions were pooled and passed through a Sephadex G-50 column from Boehringer-Mannheim (cat#100616) in 10 mM Tris-Cl pH 7.5, 1 mM

EDTA, and 0.1% SDS. Several dilutions of the dig-labeled probe were blotted onto a membrane and compared to standard amounts of a dig-labeled control DNA.

Hybridization of plaque-lifted membranes: Membranes were pre-hybridized in 50% formamide, 5X SSC, 0.1% N-lauryl sarcosine, 0.02% SDS, and 5% blocking reagent supplied by Boehringer-Mannheim (cat#1096176) and incubated at 42°C for 4 hours. The pre-hybridization solution was discarded and replaced with identical fresh hybridization solution that contained 2 μg of the dig-labeled mouse APP exon 16 probe that was boiled for 10 minutes and chilled on ice. Membranes were hybridized over a two-day period at 42°C. All incubations (and heated-washes) were performed in the Stovall "Belling Dancing" water bath. The probe/hybridization solution was removed and saved for subsequent screenings. Membranes were washed four times in 2X SSC, 0.1% SDS at room temperature for 5 minutes. Subsequent washings were as follows: two washes of 30 minutes at 65°C in 0.5X SSC, 0.1% SDS; two washes for 30 minutes at 65°C in 0.2X SSC, 0.1% SDS; ten minutes at 65°C in 0.2X SSC; and ten minutes at room temperature in 0.2X SSC.

Digoxigenin detection assay: The remaining protocol is taken from the Boehringer-Mannheim "DIG Nucleic Acid Detection Kit" (cat#1175041). Membranes were rinsed once for 2 minutes at room temperature in Genius 1 buffer (100 mM Tris-Cl, pH 7.5, 150 mM NaCl) and blocked for 1 hour at room temperature in Genius 2 buffer (2% w/v blocking agent in Genius 1 buffer). Membranes were incubated with 150 μunits/ml of polyclonal sheep anti-digoxigenin alkaline phosphatase conjugated antibody in Genius 2 buffer for 30 minutes at room temperature. Two washes were done for 15 minutes each at room temperature in Genius 1 buffer and once for 2 minutes in AP

9.5 buffer (100 mM Tris-Cl pH 9.5, 100 mM NaCl, 50 mM $MgCl_2$). Membranes were processed in Lumi-Phos 530 (Boehringer-Mannheim cat#1275470) and placed in the dark for 16 hours then exposed to film for 20 minutes.

Positive plaques were picked and placed into 1 ml SM buffer (5.8 g NaCl, 2.0 g $MgSO_4 \cdot 7H_2O$, 50 ml 1 M Tris-HCl pH 7.5 to a total volume of one liter) to diffuse and stored at 4°C. These plaques were screened in a PCR assay using the identical oligonucleotide pair that was used to generate the probe (assay-15 μl phage stock and 35 μl water were heated to 95°C for 20 minutes into which was added 10 μl 10X PCR buffer, 3 ml 100 ng/ml oligo mix of KC65 and KC66, 10 μl 2 mM dNTP mix, 5 units AmpliTaq, and 1 unit Perfect Match Polymerase Enhancer (Stratagene cat# 600129) to a total volume of 100 μl).

Secondary membrane screenings on 4 isolates were performed using the digoxigenin-mouse APP exon 16 probe previously made. Two positive phage plaques were grown (protocol taken from BioTechniques 7:21-23) to obtain enough DNA for further analysis.

A 15 Kb sequence containing the mouse APP exon 16 was sub-cloned into pBluescript IISK+ (Stratagene cat#212205) at the NotI site (designated as plasmid 35A) using standard cloning procedures. Southern analysis using a 32p-labeled mouse APP exon 16 probe revealed a 5 Kb NcoI fragment which became the backbone into which our human APP cDNAs were fused.

<u>Southern analysis</u>: Six separate reactions containing 1 μg of plasmid 35A were digested with 10 units each of restriction enzymes ApaI (cat#114S), ApaI/BglII (cat#144L), NcoI (cat#193L), NcoI/BglII, XbaI (cat#145S), XbaI/BglII (supplied by New England Biolabs) in their respective buffers (total volume of 30 μl) at their respective incubation temperatures for 3 hours. One-half of the digestion reactions was loaded onto an 0.8% agarose

(Bio-Rad cat#162-0133) gel in 1X TBE buffer. The gel was run at 20 volts overnight at room temperature. After photographing the gel, it was prepared for transferring to a nylon membrane. The gel was soaked in 0.25N HCl, rocked gently, for 15 minutes, rinsed well with water then soaked in 0.4N NaOH, rocked gently, for 20 minutes. A 3MM paper wick transfer was assembled using Amersham Hybond-N+ nylon membrane in 0.4N NaOH buffer overnight at room temperature. The membrane was rinsed in 5X SSC for 10 minutes at room temperature and UV cross-linked in a Stratalinker (Stratagene cat. #400071) using 1.2×10^5 mjoules for 30 seconds. The membrane was hybridized in 50% deionized formamide, 5X SSC, 0.1% N-lauryl sarcosine, 0.02% SDS, and 5% blocking agent (Boehringer-Mannheim) at 42°C, rocked gently, and incubated overnight. This solution was removed and replaced with the previously made mouse APP exon 16 digoxigenin-labeled probe (denatured) in fresh hybridization buffer and incubated at 42°C, rocked gently, for overnight. All subsequent washes, blocking, and antibody binding was identical to the protocol stated previously as digoxigenin detection assay.

Construction of the targeting vectors

Subcloning mouse exon 16 locus: The 5 Kb NcoI fragment containing the mouse APP exon 16 sequence was cloned into pBluescript IISK+ at an engineered NcoI site to generate pMTI2396 (Figure 2; see below). The 3 Kb 5'-flanking NcoI fragment and 2 Kb 3'-flanking NcoI fragments from p35A were also cloned into pBluescript IISK+ at the engineered NcoI site to generate pRA3 and pN2C4, respectively (Figures 3 and 4; see below). The pBluescript vector (1 μg) was digested with 20 units of XbaI in buffer 2 (NEB) and incubated for 2 hours at 37°C. Ten units of calf intestine alkaline phosphatase (CIP from Boehringer-Mannheim cat#713023) were added to the reaction and incubated for 1 hour at 37°C to dephosphorylate the 5' ends. To

500 ng of the 5Kb NcoI fragment was added 6.2 pmol of phosphorylated, annealed adapter KC95/96 (5'CTAGACACTC3') using 400 units of T4 DNA Ligase (NEB cat#202L) in its appropriate buffer (50 mM Tris-Cl pH 7.8, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP 25 mg/ml BSA) at 25°C for a 5-hour incubation. This reaction was digested with 20 units of XbaI and adjusted the buffer concentration to 50 mM NaCl and incubated for 1 hour at 37°C. The enzyme was heat inactivated at 65°C for 20 minutes. The DNA was removed from the residing enzymes using Strataclean resin (Stratagene cat#400714). To the 25 μl enzyme digestion reaction was added 5 μl of Strataclean resin, vortexed for 15 seconds and set at room temperature for 1 minute. It was then spun in an Eppendorf microcentrifuge 5415C at 14000xg for 1 minute. The supernatant was transferred to a clean tube and the procedure was repeated once. Dephosphorylated XbaI-linearized pBluescript, 50 ng, was combined with 500 ng of the phosphorylated 5Kb NcoI-adapter fragment in a standard ligation reaction and incubated at 14°C for overnight. The ligase was heat inactivated at 70°C for 10 minutes and one-tenth of the reaction was transformed into Epicurian.coli XL-1 blue cells (Stratagene cat#200236) using the protocol provided. The resulting construct having mouse genomic sequences for the 3' end of intron 15-exon 16-5' end of intron 16 was then referred to as p2396 (Figure 2). The BglII site within exon 16 is the point at which the human cDNA sequence was fused.

<u>Subcloning of NcoI fragments proximal to the Exon 16 targeting site</u>: The 3 Kb intron 15 fragment and the 2 Kb intron 16 fragment were generated by a NcoI digestion on the template plasmid 35A (see Figure 1). The 3Kb and 2 Kb NcoI-NcoI fragments were then subcloned into the Bluescript (NcoI) vector (see above). The resulting plasmids were named pRA3 (3 Kb fragment; Figure 3) and pN2C4 (2Kb fragment; Figure 4). They were expanded and the 2 and 3 Kb fragments themselves isolated by Geneclean

(Bio 101). These isolated fragments were then used as probes in Southern blot paradigms.

Generation of cloning sites around the neomycin resistance gene: As an integral part of our targeting vector construct, we cloned the neomycin resistance gene (pPol2longneobpA provided by Ann Davis) downstream of our human APP cDNA sequence. The neomycin resistance gene (contained within a pBluescript KS+ vector) was under transcriptional regulation of the DNA polymerase II promoter sequence (long version) and the bovine growth hormone (BGH) polyadenylation sequences. Sequences composed of different restriction sites had to be cloned onto both the 5' and 3' ends of this gene construct. The plasmid, 2 μg, was linearized with SalI, ligated to 45 pmol of annealed SalI-AflII-EcoRV-NcoI-MluI adapter (5'TCGACGACTTAAGTTGATATCCACCATGGTGACGCGTT3') using 400 units of T4 DNA Ligase in its appropriate buffer at 14°C in an overnight incubation. This reaction was digested with EcoRV (cat#195S) and ligated to close. This plasmid, now referred to as p2395, was digested with XhoI to linearize it at the 3' end of the BGH sequence. Ligated to this XhoI site was an XhoI-BglII-StuI adapter (5'TCGAGTGAGATCTTAAGGCCTGG3'). The ligase was removed from the reaction using the Wizard DNA clean up system (Promega cat#A7280) following the directions supplied in the kit. The linearized plasmid-adapter DNA (approx. 5 μg) was digested with 30 units each, in one 50 μl reaction, of StuI (cat#187L)/EcoRV in restriction enzyme buffer 2 (from NEB) at 37°C for 3 hours. The digest reaction was run through a 0.8% low melt agarose (FMC cat#50112) gel in 0.5X TAE buffer (20 mM Tris acetate, 0.5 mM EDTA) at 75V for 2 hours at room temperature. The 1800 bp band containing the promoter/neomycin/polyA sequences was excised from the gel and extracted from the agarose using the

Wizard DNA clean up system. This fragment was ligated to the human APP cDNA-adapter generated through the follow process.

Generation of human APP cDNA's with either the Swedish-FAD, London-FAD, Swedish/London-FAD, or Swedish-FAD, APP713 mutations fused with human APP genomic sequences containing APP polyadenylation signals: Plasmid pMTI-2385-Swedish (not shown) possesses the entire human APP 695 cDNA fused with human APP cloned into pBluescript II SK+. The plasmid pMTI 2398 was derived from pMTI2385. The strategy for its creation involved the extensive use of a cDNA-genomic hybrid plasmid, pMTI2339. pMTI2385-Swedish was assembled in a four-part ligation with the following components; an ~1861 bps. XmaI-BglII fragment from pMTI2339, a ~2008 bps. SpeI-SalI fragment from pMTI2339, a ~589 bps. fragment from FAD clone #5 (contains Swedish-FAD mutation) generated by Dr. Gerhard Konig, and a pBSSK(+)II vector opened up with XmaI and SalI. The ligation was done according to standard protocols with the insert fragments being in equal molar ratios and there being a 3:1 ratio of total insert to vector. Ligation mixtures were transformed in XL-1 Blue competent cells (Stratagene) and mini-preps analyzed by an initial digestion of XmaI-SalI. Two putative clones were further characterized with BglII-SpeI, XmaI-BglII, SpeI-SalI, EcoRI, HincII, and PvuII. Two clones, #4 and #5 gave the expected results. These were grown up and sequenced confirmed.

Plasmid pMTI-2453 was derived from pMTI-2385-London. pMTI-2385-London was assembled in a four-part ligation with the following components: a ~1.7 Kb Xma I-SacI fragment from pMTI2385-Swedish, a ~350 bp SacI-StyI fragment from pMTI-104 (contains London-FAD mutation; obtained from Paul Fracasso), a ~2.5 Kb StyI-SalI fragment from pMTI2385-Swedish, and a ~2.7 kb SalI-XmaI fragment from pMTI2385-Swedish.

Plasmid pMTI-2454 was derived from pMTI-2385-Swedish/London. Swedish/London was assembled in a four-part ligation with the following components: a ~1.9 Kb XmaI-EcoRI fragment from pMTI2385-Swedish, a ~700 bp EcoRI-ClaI fragment from pMTI-2385-London, a ~1.9 Kb ClaI-SalI fragment from pMTI2385-Swedish, and a ~2.7 Kb SalI-XmaI fragment from pMTI2385-Swedish.

Plasmid pMTI-2455 was derived from pMTI-2385-Swedish APP713. pMTI-2385-Swedish APP713 was assembled in multi-step process using PCR mutagenesis to introduce the APP713 stop mutation into proximity with the Swedish-FAD mutation. First, a ~560 bp EcoRI-SpeI fragment from pMTI2385-Swedish was ligated with the 2.9 Kb EcoRI-SpeI fragment from Bluescript KS+II (Stratagene) to generate pMTI-X. A ~400 bp fragment containing the APP 713 stop mutation was generated by PCR using APP cDNA as template and oligonucleotides RA39 (CCATCGATGGATCAGTTACGGAAACGATGCTCTCATGC) and RA40 (CCATCGATGGCCAAGGTGATGACGATCACTGTGGATCCCTACGCT ATGACAACACCGC) (Figure 9). The ~400 bp PCR fragment was digested with ClaI and StyI and ligated into the ~3.3 Kb ClaI-StyI fragment from pMTI-X to generate pMTI-Y. pMTI-2385-Swedish APP713 was assembled in a four-part ligation with the following components: a ~560 bp EcoRI-SpeI fragment from pMTI-Y, a ~1.9 Kb XmaI-EcoRI fragment from pMTI2385-Swedish, a ~2 Kb Kb SpeI-SalI fragment from pMTI2339, and a ~2.8 Kb fragment from Bluescript SK+II.

<u>Generation of the human APP "Swedish" FAD mutation cDNA-neomycin sequences to fuse to the mouse APP exon 16 DNA</u>: Four μg of plasmid pMTI-2385B was digested with 20 units of restriction enzyme SalI (cat#138L) in its ideal buffer for 2 hours at 37°C. The reaction was run through an 0.8% agarose gel in 0.5X TAE buffer at 120 volts for 1.5 hours at

room temperature. The linearized DNA band was excised and isolated away from the agarose using the Qiaex DNA Gel Extraction kit and the protocol provided (Qiagen Cat#20021). One μg of SalI-linearized p2385B was ligated to 45 pmol of annealed SalI-AflII-EcoRV-NcoI-MluI adapter (mentioned previously) in a standard ligation reaction. One-tenth of the ligation reaction was used to transform E.coli XL-1 blue cells. This p2385B-adaptor construct, 18 μg, was linearized with 60 units of EcoRV in a standard digestion reaction. Into this was added 15 units of calf intestine alkaline phosphatase and incubated at 37°C for 1 hour to dephosphorylate the 5' ends of the DNA. The reaction was stopped with EDTA at a final concentration of 5 mM and heat inactivated at 75°C for 10 minutes. The dephosphorylated plasmid was gel isolated and 1 μg was ligated to 400 ng of the 1800 bp neomycin fragment with EcoRV 5' and StuI 3' ends (mentioned in the section "Generation of cloning sites around the neomycin resistance gene"). One-tenth of the ligation reaction was used to transform E. coli XL-1 blue cells following the protocol provided by the supplier. Correct orientation constructs had the neomycin fragment (5' EcoRV site) placed immediately downstream of the human APP cDNA polyA sequences (3' EcoRV site), this construct was designated p2397+A (not shown).

Construction of the completed targeting vector containing the human APP "Swedish" FAD mutation: The 5 Kb mouse APP exon 16 containing DNA, p2396 (12 μg), was digested with 50 units of BglII in buffer 3 for 3 hours at 37°C. To 6 μg of the digest was added 10 units of CIP and incubated at 37°C for 1 hour. The reaction was stopped as mentioned above and the DNA was gel isolated using Gelase (Epicentre cat#G09100) and following the supplied protocol. The 4.5 Kb BglII fragment containing the human APP cDNA-neomycin fusion was released from p2397+A by digesting 12 μg of DNA with 28 units of NruI (cat#192L) in its ideal buffer at

37°C for 3 hours. After being confirmed of its linearization 50 units of BglII were added for an additional 2-hour incubation at 37°C. The 4.5 Kb fragment was gel isolated using Gelase and then ligated (300 ng) to the dephosphorylated p2396-BglII linearized DNA (100 ng) in a standard ligation reaction and subsequently transformed into E.coli XL-1 blue cells. The resulting plasmid with the mouse APP exon 16 fused to the human APP cDNA at exon 16 (BglII site) was designated as p2398 (Figure 5 and Figure 18).

E.2.7 Cloning of the HSV thimidine kinase (TK) gene into the targeting vector: The HSV thimidine kinase gene (from pAD7) was provided by Ann Davis. Unique restriction sites had to be engineered with the TK gene to provide linearizing access in the completed targeting vector. A 3Kb BamHI-ClaI fragment containing the murine phosphoglycerate kinase (PGK) promoter regulating the TK gene with the BGH polyadenylation sequences was isolated away from vector sequences and sub-cloned into pBluescript II SK+ at its respective sites. Twenty μg of this new TK plasmid, pCB11, was digested with 60 units of SalI in its unique buffer and incubated overnight at 37°C. The enzyme was heat inactivated at 65°C for 20 minutes and then 10 units of CIP was added for 1 hour at 37°C. The phosphatase was heat inactivated at 75°C for 10 minutes. The linearized DNA band was excised and isolated away from the agarose using the Qiaex DNA Gel Extraction kit and the protocol provided as stated above. In a standard ligation reaction, 45 ng of SalI-linearized vector was added to 15 pmol of annealed SalI-AscI-PmeI-NotI-AscI-PmeI-SalI adapter (5'TCGACAAGGCGCGCCGTTTAAACAAGCGGCCGCTTGGCGCGCCT TTTGTTTAAACTTG3') and incubated overnight at 14°C. This TK plasmid containing the restriction sites PmeI and AscI was designated as pXII28N. Five μg of pXII28N was digested with 20 units of NotI (cat#189L) and 15 units of PvuI (cat#150L) in NotI buffer (NEB) and 0.1 mg/ml BSA at 37°C

for overnight. The 3Kb NotI TK band was excised and isolated away from the agarose using the Qiaex DNA Gel Extraction kit and the protocol provided. Five hundred ng of this TK fragment was ligated to 50 ng of NotI linearized p2398 (vector containing the APP/neo sequences fused to the mouse APP exon 16 sequences) in a standard ligation reaction and incubated overnight at 14°C. The resulting targeting vector, p2399 (350 μg), was linearized with 320 units of PmeI (cat#560L) in buffer 4 (NEB) and 0.1 mg/ml BSA and incubated overnight at 37°C. Protein was removed by adding sodium acetate pH 5.2 to 0.3 M and extracting twice with Tris-Cl buffered phenol and extracting once with chloroform and ethanol precipitating at -20°C for overnight.

<u>Construction of the completed targeting vectors containing the human APP London-FAD, Swedish/London-FAD, and "Swedish-FAD APP713 mutation</u>: These three targeting vectors were constructed by ligating four separate fragments with one of these fragments containing one of the FAD mutations. The seminal targeting vector construct, p2398, was digested in three independent reactions to obtain three of the specific fragments. (1.) Five μg of p2398 were digested with 15 units of AflII (cat #520S) in buffer 2 with 0.1 mg/ml BSA at 37°C for overnight. To this reaction was added enough buffer 3 to adjust the concentration to 100 mM NaCl and 30 units of NotI and incubated at 37°C for 3 hours. (2.) Twenty μg of p2398 were digested with 24 units of BglII, 20 units of NotI, and 0.1 mg/ml BSA in buffer 3 at 37°C for overnight. (3.) Twenty μg of p2398 were digested with 20 units of AflII, 10 units of ClaI (cat#197L), and 0.1 mg/ml BSA in buffer 4 at 37°C for overnight. All three digestion reactions were run on 0.8% low melt agarose gels in 0.5X TAE buffer at 70 volts for 3 hours. From digestion reaction (1.) an 8Kb fragment containing the neomycin-murine APP intron 16-pBluescript sequences was excised, from reaction (2.) a 2Kb

fragment containing murine APP intron 15-exon 16 sequences was excised, from reaction (3.) a 2Kb fragment containing human cDNA/polyA sequences was excised and all the fragments were isolated away from the agarose using the Qiaex Gel Extraction kit. The last fragments to isolate were the three 700bp human APP FAD containing fragments. Twenty-five μg of each of the pMTI-2385 London (not shown), pMTI-2385 Swedish/London (not shown), and pMTI-2385 Swedish-FAD 713 (not shown), vectors were digested with 24 units of BglII, 15 units of ClaI, and 0.1 mg/ml BSA in buffer 4 at 37°C for overnight. The 700 bp bands from these digestions were isolated away from the agarose using the identical protocol as above. A four-part standard ligation reaction was combined using 25 ng of the 8 Kb AflII/NotI fragment, 250 ng of the 2 Kb AflII/ClaI fragment, 300 ng of the 700 bp BglII/ClaI fragment, and 250 ng of the 2 Kb NotI/BglII fragment and incubated at 14°C for 24 hours. One-sixth of the ligation reaction was used to transform E.coli XL-1 blue cells in a standard protocol. The resulting constructs were designated as p2450 (London-FAD), p2451 (Swedish/London-FAD), and p2452 (Swedish-FAD APP713)(not shown). The final step for each individual plasmid was to clone the TK gene fragment with NotI ends into it. Five μg of each plasmid, p2450, p2451, p2452 were digested with 20 units of NotI in buffer 3 at 37°C for 3 hours. To dephosphorylate the vector, 10 units of CIP were added to the digestion reaction and incubated at 37°C for 1 hour. The phosphatase was heat inactivated at 75°C for 10 minutes. The linearized DNA band was excised and isolated away from the agarose using the Qiaex DNA Gel Extraction kit and the protocol provided as stated above. Fifty ng of each dephosphorylated vector was ligated to 300 ng of the 3 Kb NotI TK gene fragment in a standard ligation reaction. The resulting plasmids were designated as p2453 (London-FAD; Figure 6: Figure 19), p2454 (Swedish/London-FAD; Figure 7; Figure 20), and p2455 (Swedish-FAD APP713; Figure 8: Figure 21). Each of these three targeting vectors

(500 μg) were linearized with 500 units of AscI in buffer 4 at 37°C for overnight. The DNAs were cleaned away from the enzymes by phenol/chloroform extractions as stated in the section "Cloning of the HSV thimidine kinase (TK) gene into the targeting vector". Linearized plasmids were electroporated into ES cells.

miniSouthern-blot analyses

DNA sample preparation: Potential clones were grown in a 96 well plate format. Samples were lysed with the addition of 50 μl of Lysis Buffer [10 mM Tris pH 7.5, 10 mM EDTA pH 8.0, 10 mM NaCl, 0.5% Sarcosyl, and 1 mg/ml Proteinase K (added fresh)] per well and incubated overnight at 65°C in a humidified chamber. The DNA is precipitated by the addition of 100 μl of 75 mM NaCl in ethanol followed by incubation at room temperature for 15-30 minutes. The DNA is then washed 3x with 150 μl of 70% ethanol added drop by drop to each well. After the final wash, the plate is inverted and allowed to air-dry for 5 - 10 minutes. While the plate is drying, the Restriction Enzyme Cocktail (1x Restriction Buffer specified for the enzyme being used, 1 mM Spermidine, 100 μg/ml Bovine Serum Albumin, and 10 - 20 units of enzyme) is prepared. 30 μl of this cocktail is then added to each well. Incubate overnight at the restriction enzyme's required temperature in a humidified chamber. The next day add 4-5 ml of loading dye (10 mM Tris-HCl pH 6.0, 0.25% bromophenol blue, 0.25% xylene cyanol FF, 15% Ficoll (Type 400; Pharmacia) in water, and 30 mM EDTA.) and store at -20°C.

Agarose Gel Electrophoresis: A large gel tray (Owl Scientific) is prepared with three 36-teeth combs (evenly distributed along the length of the tray) and 400 ml of molten agarose (FMC). This size of gel will accommodate one 96-well mini-Southern digest plate. The samples were

electrophoresed for approximately three hours at 120 V. After the electrophoresis was complete, the gel was denatured in 0.25 M HCl (2 x 7 minutes at room temperature) and then equilibrated in 0.4 N NaOH (1 x 20 minutes at room temperature). An overnight alkaline capillary transfer is set up in 0.4 N NaOH with Gene Screen Plus (DuPont NEN). The next day the membrane was neutralized in 2x SSC for 5-10 minutes and then UV cross-linked (Stratagene). The membrane is then stored dry until hybridization. Prehybridization was carried out in 1 M NaCl (Gibco BRL), 10% Dextran Sulphate (Pharmacia), 1% SDS (Gibco BRL), and 200 μg/ml salmon sperm DNA(Stratagene) for at least one hour at 65°C in a Robbins Hybridization Oven. The probe of interest was then labeled according to the standard protocol contained in the Prime-It II random prime kit (Stratagene). The specific activity of the probe was approximately 1 x 109 dpm/μg. It was then added directly to the prehybridization mixture at a concentration ~1 x 106 dpm/ml. The filter(s) were then hybridized for 16 hours at 65°C in the hybridization oven. The initial post - hybridization wash was carried out for 5-10 minutes at room temperature in 2x SSC (3 M NaCl, 0.3 M Sodium Citrate Dihydrate), 1% SDS. A stringent wash was then performed in 1x SSC, 0.1% SDS at 65°C for 30 minutes. The filter was then placed into a seal-a-meal bag and placed into a Fuji Phosphoimager for interpretation.

Confirmatory Southern Blots

Preparation of High Molecular Weight DNA from Cells: In order to confirm targeted clones identified in the mini - Southern paradigm, cell pellets expanded from these clones are analyzed for accurate recombination events at both the 5' and 3' ends of the targeting vector. 1 ml of Cell Lysis Buffer (100 mM NaCl, 50 mM Tris pH 7.5, 10 mM EDTA pH 8.0, and 0.5% SDS) and 20 ml of freshly prepared 40 mg/ml Proteinase K (Boehringer-Mannheim) was added to each cell pellet. The tubes were

rocked overnight at 65°C. The next day, an equal volume amount of isopropanol was added and the tube inverted several times to precipitate the DNA. The DNA was then spooled onto a flame sealed micropipette and rinsed once in 70% ethanol, once in 100% ethanol, and then air dried. The pipette was broken off into a sterile Eppendorf tube and the DNA dissolved in 200 μl of sterile TE overnight at room temperature. The DNA is then stored at 4°C until restriction enzyme analysis.

Agarose Gel Electrophoresis: Restriction enzyme digested DNA is gel analyzed as described above in the mini-Southern methods except the number and sizes of the combs vary. Denaturation, renaturation, and capillary transfer were performed as described previously. Probes of interest were also labeled in the same manner as described above. Interpretation of results were facilitated by phosphoimaging as previously described.

Gene-targeting in ES cells

Culture of ES cells: Procedures were performed essentially as describe in E. J. Roberstion (Robertson, 1987) . ES cell were propagated using Mitomycin C treated SNL76/7 STO feeder cells (cell line obtained from A. Bradley) and modified DMEM culture media (supplemented with 15% FCS,1X GPS,1X BME).

Electroporation of ES cells: DNA was linearized with the appropriate restriction enzyme then extracted with an equal volume of phenol/chloroform and once with an equal volume of chloroform and precipitated with 2.4 volumes of ethanol. The DNA was resuspended at 1 mg/ml in sterile 0.1X TE (25 ml of DNA per electroporation). Embryonic stem cells (80% confluent) were passaged 1:2 the day before electroporation. Cells to be electroporated were fed 4 hours before harvesting. The cells were trypsinized and resuspend in media (cells from 2 x 10 cm plates can be combined in a total volume of 10 ml in a 15 ml tube). The cells were pelleted and resuspend in 10 ml PBS at a density of 11 x 10^6 cells/ml. The appropriate amounts of DNA and cells were mixed together in a 15 ml tube (25 ml of DNA and 0.9 ml of cells for each electroporation) and allowed to sit at room temperature for 5 minutes. The cell/DNA mixture (0.9 ml) was transferred to electroporation cuvettes and an electrical current was passed through the solution (using Biorad GenePulser at 230V and 500 mF). The cells were then transferred to culture plates with feeder cells (up to 2 x 10^7 cells/100 mm plate or 6 x 10^6 cells/60 mm plate). After 24 hours of culture in modified DMEM the cells were cultured in DMEM selection containing G418 and 0.2 mM FIAU. Resistant colonies may be picked as early as 8 days, are best around 10-11 days, but may be recovered up to 18-21 days after the electroporation. Picked colonies are transferred to 96 well plates with feeders

cells and screened for gene-targeting events by mini-Southtern-blot analysis (see below).

Production of chimeric mice: Procedures were performed essentially as described by A. Bradley (Bradley, 1987). Host 3.5 day blastocysts were derived from timed matings of C57BL/6 mice and cultured in M16 media. Approximately 14 targeted ES cells were injected into each blastomere. Surviving blastocysts were then surgically reimplanted (approximately 12 per animal) into pseudopregnant ICR female mice essentially as described (A. Bradley). Chimeric mice were born about 17 days after implantation.

Genotype analyses of transgenic mice

Identification of mice possessing the targeted human APP cDNA by PCR screening: When mice were older than 2 weeks of age their tails were biopsies to obtain genomic DNA for analysis. One centimeter pieces of tail were prepared using the QIAamp Tissue Kit (Qiagen cat# 29304) and following the protocol provided. Genomic DNA was eluted in 150 μl of 10 mM Tris-Cl pH 9 and used in two independent PCR assays; (1) to determine the endogenous mouse APP allele that remained intact: total reaction volume of 50 μl - 5 μl of genomic tail DNA (approximately 1 μg), 5 μl of 10X buffer 8 (Stratagene cat#200430), 5 μl of 2 mM dNTP mix, 200 ng of oligonucleotide KC125 (5'ACTTTGTGTTTGACGC3'), 200 ng of oligonucleotide KC132 (5'CAGTTTTTGATGGCGG3'), 1 unit of Perfect Match Polymerase Enhancer, 2.5 units of AmpliTaq and 100 ng each of oligonucleotides 6&7 and (2) to determine the targeted mouse APP allele: total reaction volume of 50 μl - 5 μl of genomic tail DNA (approximately 1 μg), 5 μl of 10X buffer 8 (Stratagene cat#200430), 5 μl of 2 mM dNTP mix, 200 ng of oligonucleotide KC125 (5'ACTTTGTGTTTGACGC3'), 200 ng of oligonucleotide KC131 (5'GATGATGAACTTCATATCCTG3'), 1 unit of

Perfect Match Polymerase Enhancer, 2.5 units of AmpliTaq and 100 ng each of oligonucleotides 6&7. The reactions were run in a Perkin Elmer turbo 2400 thermal cycler. The parameters of the run were as follows: one cycle at 94°C for one minute, 30 cycles at 94°C for 30 seconds-56°C for 50 seconds-70°C for two minutes, maintain at 10°C indefinitely. Oligonucleotides 6 (5'CCTCGGCCTTTGGTGTGTGTTTTATGACATGACCCCCTTGA) & 7 (5'CACCCTGTTGTCAATGCCTCTGGGTTTCCGCCAGTTTCG3') are homologous to mouse ribosomal protein L32 sequences within intron 2 and exon 3, respectively, and used as an internal DNA control signal. One-fifth of each PCR reaction was run on a 6% polyacrylamide gel (Novex cat#EC6265) in 1X TBE (89 mM Tris borate, 2 mM EDTA) buffer at 125 volts for 35 minutes and stained in 1 mg/ml EtBr for 15 minutes and photographed.

<u>RNA analyses</u>

<u>RNA isolation</u>: Total brains were dissected and flash frozen on dry ice from two negative litter mates, two heterozygous targeted mice, and two homozygous targeted mice. In addition, kidneys and tails were also removed from these mice and flash frozen. The brains were divided in half, one for the RNA analysis and the other for protein analysis. To one-half of each brain was added 5 ml RNAzolB (Tel-Test, Inc. cat#CS-105) and the tissues were homogenized using a Brinkman Polytron at medium speed for 20 seconds. Chloroform, 500 μl, was added to the homogenized tissue and shaken well for 10 seconds and incubated on ice for 15 minutes. The samples were spun in a tabletop Sorvall centrifuge at 1500Xg for 20 minutes at 4°C. The aqueous phase was removed and added to an equal volume of isopropanol, mixed, and incubated on ice for 15 minutes. The samples were spun in a Sorvall RC-5B centrifuge with an SS-34 rotor at 7500Xg at 4°C for

25 minutes. The supernatants were removed and the pellets were rinsed twice in cold 70% EtOH and air dried. The total RNA pellets were resuspended in 500 ml H_2O and incubated at 65°C for 10 minutes to more easily get the RNA into suspension. These RNA samples were used to obtain polyadenylation specific mRNA using the PolyAtract mRNA isolation system III kit (Promega Z5300). The protocol followed was provided by the supplier and yields ranged from 3 to 6 μg of mRNA.

Northern blot analyses: These samples were then used in a Northern blot to see the sizes of these targeted hybrid APP transcripts. The RNA was run on a 1.2% agarose (FMC cat# 50072) , 2.2 M formaldehyde gel prepared as follows: 0.6g agarose in 36 ml H_2O were melted in a microwave and placed at 60°C. When the gel cooled to 60°C , 5 ml of 10X MOPS (0.4 M MOPS (Sigma MESA M-5755) pH7, 0.1 M sodium acetate, 10 mM EDTA pH8) running buffer and 9 ml of 37% formaldehyde (pH >4) were added, mixed and left at 45°C until ready to pour. The RNA samples were prepared in a total volume of 30 μl - 3 μl 10X MOPS buffer, 5.25 μl 37% formaldehyde, 15 μl formamide, and 6.75 μl mRNA (0.5 μg) were mixed well and incubate at 55°C for 15 minutes. To this was added 6 ml formaldehyde loading buffer (1 mM EDTA pH8, 0.25% bromophenol blue, 0.25% xylene cyanol, 50% glycerol) and 1 ml 1 mg/ml EtBr. The samples were loaded into the gel and run at 5V/cm (55-75V) for 3hr in 1X MOPS buffer. The gel was rinsed in H_2O several times and soaked in 0.05N NaOH for 30 minutes under gentle shaking. The gel was then equilibrated twice for 15 minutes in 20X SSC and transferred by wick assembly for 16 hours in 20X SSC. The membrane used for transferring was Hybond-N+ (Amersham cat#RPN2020B) which is a 0.45 micron nylon membrane. After transfer was completed the membrane was rinsed in 2X SSC for 10 minutes and UV cross-linked in a Stratalinker mentioned earlier. The membrane was pre-hybridized

in 10 ml 0.5 M sodium phosphate pH7, 1% BSA, and 7% SDS for 4 hours at 65°C in a Robbins Hybridization Oven (model 400). This solution was removed and replaced with fresh hybridization solution and 6x107 counts of denatured APP probe and 8x105 counts of denatured mouse beta-actin probe and hybridized overnight at 65°C. The membrane was washed in 2X SSPE, 0.1% SDS at 25°C for 10 minutes, twice, and then washed in 1X SSPE, 0.1% SDS (pre-warmed) at 65°C for 15 minutes. The membrane was exposed to a phosphoimaging screen for 24 hours and developed.

<u>Probes for Northern blot</u>: Both the APP probe (homologous to the murine and human sequences) and the murine beta-actin probe were prepared in identical protocols. The APP DNA used to make the probe was an NruI/XhoI 900bp fragment from p2385B. The murine beta-actin 430bp DNA used for the probe came from a PCR reaction where the exon 3 of B-actin was amplified using these two oligonucleotides: KC137 (5'GTTTGAGACCTTCAACACCC3') and KC138 (5'GAAGGAAGGCTGGAAAAGAGCC3'). The probes were labeled using the Prime It II kit (Stratagene cat#300385) and following the protocol provided. After the reactions were stopped they were put over a G-50 spin column (5'-3' cat# 5303-633329) to remove the un-incorporated nucleotides.

To increase the level of APP-specific mRNA from the polyA selected RNA, the samples were annealed to an APP specific oligonucleotide (RA49-5'CGATGGGTAGTGAAGCA3')) that was homologous to both the murine and human sequences approximately 40nt 3' of the stop codon. The assay was performed using the Superscript II RT-PCR kit (Gibco/BRL cat# 18089-011). In a reaction volume of 14 μl was combined 0.1-0.15 μg polyA mRNA and 600 ng RA49 and incubated at 70°C for 10 minutes and 4°C for 10 minutes. To this was added 2 μl 10X synthesis buffer, 1 μl 10X dNTP mix, 2 μl 0.1 M DTT, and 200 units of Superscript II reverse transcriptase (all

supplied by the kit) and the incubations continued at 25°C for 10 minutes, 42°C for 50 minutes, 70°C for 15 minutes, and 4°C for 10 minutes. The reactions were then treated with 2 units of RNaseH for 20 minutes at 37°C and then placed on ice. After the RNA was removed from the cDNA the next step was the amplification reaction: 20 μl of cDNA reaction mix, 8 μl 10X synthesis buffer, 300 ng of oligonucleotide KC56 (5'GTGAAGATGGATGCAGAATTC3'), 300 ng of oligonucleotide KC56 Swedish (5'GTGAATCTAGATGCAGAATTC3'), 600 ng of oligonucleotide RA49, and 5 units of AmpliTaq in a total volume of 100 ml. The amplification was run in the Perkin Elmer turbo 2400 using the same parameters as stated in "Identification of mice possessing the targeted human APP cDNA by PCR screening". The RT-PCR reactions were subjected to restriction enzyme digestions taking advantage of the restriction site polymorphism between the murine and human APP sequences. One-tenth of the RT-PCR reaction was digested with 30 units of SalI and 0.1 mg/ml BSA in its ideal buffer at 37°C for 2 hours, another set was digested with 30 units of StyI in buffer 3 at 37°C for 2 hours. The digests were run out on a 4% polyacrylamide gel in 1X TBE at 150 volts for 1 hour and stained in 1 mg/ml EtBr for 15 minutes and photographed. All oligos were provided by Midland and all restriction enzymes by NEB.

Protein Analysis

Tissue Extraction: This protocol is generally used for mouse tissue with no more than several hundred mgs of tissue available, therefore all volumes must be kept to a minimum. Tissue was homogenized in 1 ml of RAB buffer (0.1 M MES pH 7.0, 0.75 M NaCl, 0.5 M $MgCl_2$, 1 mM EGTA, 1 mM DTT) containing proteinase inhibitors. The protease inhibitor cocktail contains 1x Aprotonin (0.41 trypsin inhibitor units/mg protein), 1x PMSF (2 mM in isopropanol), 1x Protease inhibitor mix (chymostatin,

leupeptin, antipain, and pepstatin) each at 50 μg/ml in DMSO, and 1 mM EDTA. A 7 ml dounce tissue grinder (Wheaton) was used for homogenization. The tissue homogenate was spun at 40K in the Beckman TL100 using the fixed angle rotor for one hour. The supernatant from this spin was saved as it contains the soluble APP. The pellet was homogenized in 1 ml of RAB plus protease inhibitors and 30% sucrose (Sigma). Spin for one hour at 40K in the Beckman TL100. This serves as a wash and demylelinating step. Discard the supernatant from this spin and homogenize the pellet in 1 ml of RIPA buffer (150 mM NaCl, 1% NP-40, 0.5% deoxycholate (Na salt), 0.1% SDS, and 50 mM Tris-Cl pH 8.0). This should contain the membrane associated form of APP. The amount of protein can then be quantitated by using the BCA Protein Assay Reagent Kit (Pierce). This quantitation allows equal amounts of total protein to be loaded on polyacrylamide gels and direct comparisons of transgenic and non-transgenic expression patterns and levels to be made.

Immunoprecipation: The final adjusted volume of the immunoprecipitation was 1 ml in RIPA buffer. The amounts of antigen and antibody to add varied from experiment to experiment depending on the concentrations of both. Antibody and antigen were incubated for two hours at 4°C while gently spinning on a rotating wheel. 50 μl of goat anti-mouse or anti-rabbit IgG bound to agarose (Sigma) was added to the antigen/antibody and incubated for another two hours at 4°C on the rotating wheel. Agarose IgG-antigen/antibody complex was rinsed by pelleting at 12,000 x g for 1 min. and then removing the supernatant. Then 500 μl of ice cold RIPA buffer was added to the pellet, resuspended, and incubated for 10 minutes on ice. The samples were then spun at 4°C. The rinses were repeated twice more, but the 10 minute incubation step was omitted. To the rinsed pellet, was added 50 μl of 1x sample buffer (Novex)plus 2 ml of beta-

mercaptoethanol (Aldrich). Samples were boiled for 10 minutes and spun for 1 minute at room temperature. The supernatant was transfered to a fresh tube and store at -20°C.

Western Blotting: Polyacrylamide gel electrophoresis (PAGE) and electroblottting were accomplished utilizing the X-Cell II Gel and Blot Module (Novex) and pre-casted polyacrylamide gels (Novex). The selection of a particular separation scheme depended on what form of the Alzheimer Precursor Protein (APP) was being examined. For C-terminal fragments 16% Tris-Tricine gels were utilized, holo APP utilized 10-20% Tris-Tricine gels, and to elucidate form differences (Kunitz vs. 695) of the holo-APP, 6% Tris-Glycine gels were used. Samples prepared as described above were loaded onto gels and electrophoresed at 120V for approximately 90 minutes. The gels were then transferred to nitrocellulose membranes (Novex) for 1-2 hours at 30V. Non-specific sites were then blocked by incubation of the filter in 5% non-fat dry milk (NFDM) for 1 hour at room temperature while gently rocking. Primary antibody was then added at a dilution of 1:500 in 5-10 ml of NFDM, added to the membrane and sealed in a seal-a-meal bag. This was incubated overnight at room temperature while gently rocking. The membrane was then rinsed for 1 hour at room temperature with several changes of 5% NFDM. A ^{35}S labeled secondary antibody (Amersham), either anti-mouse IgG or anti-goat IgG, was then added and incubated for 1 hour at room temperature while gently shaking. The membrane was then rinsed for 15-30 minutes in 5% NFDM and then equilibrated into 1x phosphate buffered saline (PBS, Gibco BRL) for 15 minutes. The filter was then dried and either placed on a phosphoimaging plate or with a piece of X-OMAT X-ray film (Kodak).

APP Antibodies: Monoclonal antibody (MAb) 4G8 (Senetek) was used for the immunoprecipitation of APP holoprotein and C-terminal fragments at a dilution of 1:100 (~10-20 μg/ml). MAb 286.8 (BRC) was used for the immunoprecipitation of APP holoprotein at a dilution of 1:100 (~10-20 μg/ml). MAb 6E10 (Senetek) was used as a detection reagent on Western blots at dilutions of 1:500. Polyclonal antibody (PAb) 369 (generously provided by Dr. Sam Gandy) was used for both the immunoprecipitation of APP holo-protein (1:100) and for a detection reagent for C-terminal fragments (1:500). MAb 22C11 (generously provided by Dr. Konrad Beyreuther) was used as a detection agent for APP holo-protein at a dilution of 1:500.

FAD-m/hAPP gene products expressed in transgenic mouse lines

Transgenic mouse lines ES5007, ES5103, ES5401, and ES5403 were generated by mutating the mouse APP gene via homologous recombination in embryonic stem (ES) cells (see below). The gene products expressed in the transgenic mouse lines are described schematically in Figure 10. m/hAPP770 represents the largest (770 amino acid residues) of the various alternative splice forms of protein expressed by each mutated mouse APP gene. m/hAPP exhibits amino acid sequence identity with mouse APP with the exception of those residues indicated by (asterisks, *). In all cases the beta-amyloid (bA4) domain (Asp672 to Thr714; 43 amino acid residues) has been "humanized" by the introduction of three amino acid substitutions (as indicated by green asterisks); Gly(676) to Arg, Phe(681) to Thr, and Arg(684) to His. Transgenic mouse line ES5007 also has the Swedish-FAD mutation [Lys,Met(670,671) to Asn,Leu] introduced into the mouse gene. Transgenic mouse lines ES5401 and ES5403 have the London-FAD mutation [Val(717) to Ilu] and transgenic line ES5103 carries both Swedish and London FAD mutations. In addition to the Swedish FAD and "human"

mutations, a transgenic mouse line ES5215 can also be produced which has a premature stop codon introduced at position 714.

Gene-Targeting Vectors

The targeting vectors were designed in such a way as to facilitate the integration of human cDNA sequences into mouse exon 16. The targeting constructs function as replacement-type vectors with both positive (neomycin resistance gene) and negative (HSV TK gene) selection genes (figure 11). To facilitate homologous recombination, a mouse genomic clone encompassing exon 16 was obtained by screening a mouse genomic lambda library. A lambda clone, "35A", was identified which contained an intact exon 16 (figure 1). Nco I fragments of mouse genomic clone 35A were subcloned into the BSII SK+ vector and the subclones (pRA3, pMTI-2396, and pN2C4) were characterized by DNA sequence and restriction enzyme analyses (see Figures 2, 3, and 4). The 5.5 Kb Nco I DNA fragment (subcloned into pMTI-2396) contains APP exon 16 and ~1.9 Kb and ~3.5 Kb from introns 15 and 16 respectively (Figure 2). The Nco I DNA fragment, containing exon 16, was the template upon which the gene-targeting vectors were constructed.

The gene-targeting vectors were designed so that mouse exon 16 gene sequences were fused (at a common Bgl II site) with human cDNA sequences which encode the remainder of exon 16 and exons 17 and 18 (figure 11). The mouse and human cDNA sequences encode the identical protein sequence with the exception of 3 amino acid differences (shown as green asterisks) which reside within the beta-amyloid domain. The mouse genomic-human cDNA fusion effectively "humanized" the beta-amyloid domain and facilitated the introduction of specific FAD mutations while leaving the remainder of mouse APP protein sequences unchanged (see Fig 13).

The human cDNA was mutagenized to encode either the "Swedish"-FAD, "London"-FAD , "Swedish"/"London"-FAD (shown here), or "Swedish"-FAD APP713 mutations (shown as black asterisks) of APP (see Fig. 10 and Table I). The mutagenesis of the "Swedish"-FAD mutation also incorporated a new Xba I restriction site. Proper RNA processing was ensured by fusing the 3'-end of the human cDNA sequence with human genomic sequences which contain transcription termination and poly-adenylation signals from the human APP gene. A neomycin gene was inserted in-between the 3'-end of the human APP polyadenylation signal and mouse APP intron 16 sequences. Targeting vector pMTI-2398 (Swedish-FAD) contained the neomycin resistance gene and not the HSV TK gene. This vector was linearized with Pme I and was used to generate transgenic mouse line ES5007. (Tables I and II).

For the remaining three targeting vectors, a HSV Tk gene was inserted into the clone in such a way that its placement was outside of the genomic domains homologous to mouse (Figure 11; as shown or in the opposite orientation; the orientation was not critical). Targeting vector pMTI-5453 encodes London-FAD m/hAPP, targeting vector pMTI5454 encodes Swedish/London-FAD m/hAPP, and targeting vector pMTI-5455 encodes Swedish-FAD m/hAPP713. These targeting vectors were linearized with Asc I and were used to generate transgenic lines. (Tables I and II).

<u>Gene-targeting in embryonic stem (ES) cells</u>

The targeting vectors were designed to function as replacement-type vectors with both positive (neomycin resistance gene) and negative (HSV TK gene) selection genes. After electroporation of the targeting vector into ES cells, G418 drug treatment selected for ES cells which had integrated the targeting vector (including the neomycin resistance gene) into the mouse genome. The majority of G418 resistant ES cell clones had targeting vector

integrated at random locations of the genome. These ES cell clones retained an intact HSV TK gene and were not desired. The clones containing random integrations could be eliminated by treatment with FIAU selection media which is toxic only to cells expressing HSV TK. If, as desired, the mouse APP gene is targeted via a double-crossover homologous recombination event, the flanking non-homologous HSV TK DNA sequences are lost (as shown in fig 12) and the ES cells are resistant to FIAU treatment.

Homologous recombination between mouse APP exon 16 locus and the gene-targeting vector fundamentally alters the manner by which the gene encodes APP (see figure 13). Normally, the beta-amyloid, transmembrane, and cytoplasmic domains of mouse APP are encoded by three separate exons. In addition, the coding region for the beta-amyloid domain resides both on exons 16 and 17. After homologous recombination with the gene targeting vector, however, mouse exon 16 gene sequences are fused with human cDNA sequences. Mouse exons 17 and 18 are now displaced down-stream from the neomycin resistance gene and are inactive. The human cDNA now functions in place of mouse exons 16, 17, and 18 to encode APP. Therefore, the beta-amyloid, transmembrane, and cytoplasmic domains of mouse APP are now encoded by human cDNA sequences.

The gene products of this new mouse genomic-human cDNA fusion are designated m/hAPP. Human cDNA sequences (exons16, 17, and 18) encode the identical protein sequence with the exception of 3 amino acid differences (shown as green asterisks) which reside within the beta-amyloid domain. The mouse genomic-human cDNA fusion effectively "humanizes" the beta-amyloid domain and facilitates the introduction of specific FAD mutations (shown as black asterisks) while leaving the remainder of mouse APP protein sequences unchanged. The human cDNA has been mutagenized to encode either the "Swedish"-FAD, "London"-FAD , "Swedish"/"London"-

FAD (shown in figure 2c), or "Swedish"-FAD APP713 mutations of APP (see also Fig. 10 and Table I).

Identification of targeted ES cell clones

After electroporation of each targeting vector (see Table I), ES cells were cultured for approximately 2 weeks in the presence of both positive (G418) and negative (FIAU) selection compounds. Four hundred G418/FIAU resistant ES cell colonies (clones) were then individually picked and cultured separately in 96 well culture dishes. The culture dishes were replica-plated, one set of copies was frozen to maintain the clones and the other replicate set was utilized for genetic analyses. From each well, DNA was extracted, digested with restriction enzyme, and analyzed by miniSouthern-blot analyses (see below). ES cell clones which appear to contain a targeted APP gene locus were thawed and expanded in culture. Gene-targeting was confirmed by Southern-blot analyses using DNA extracted from these expanded clones prior to introduction of the ES cell into the mouse germline (see below).

The mutagenesis of human cDNA's to encode the Swedish-FAD mutation also created a new Xba I (shown as X) restriction enzyme site (Figure 14). Incorporation of human FAD cDNA (shown in red) into the targeted m/hAPP gene locus thus changes the pattern of DNA fragments generated after digestion of this locus with Xba I. Using Southern-blot analyses, ES cell clones having the targeted m/hAPP gene can be distinguished from neomycin resistant ES cell clones having undesired random integrations of the targeting vector. The mouse exon 16 gene locus can be detected using a 3Kb Nco I (N) DNA fragment from intron 15 of the mouse APP gene as probe. Digestion of the mouse APP gene with Xba I generates an approximately 9 Kb DNA fragment whereas Xba I digestion of the targeted Swedish-FAD m/hAPP gene gives an approximately 5 Kb DNA

fragment when detected by Southern-blot analysis (see figure 14). This detection strategy applies to the Swedish-FAD m/hAPP, Swedish/London-FAD m/hAPP, and Swedish-FAD APP713 mutations.

Mini-Southern blot analysis identified 4 ES cell clones which appeared to contain the targeted Swedish-FAD m/h APP locus (data not shown). These clones were expanded and subsequent Southern-blot analysis demonstrated that ES cell clones A79, A80, and B12 contain the Swedish FAD APP mutation while clone A72 did not (Figure 15). DNA extracted from ES cell pellets was examined by Southern-blot analysis using the restriction enzyme Xba I as described in Figure 14. A single ~9Kb DNA fragment is observed in non-targeted ES cells whereas targeted ES cell clones exhibit both the non-targeted allele (~9Kb fragment) and the FAD mutant allele giving rise to a ~5 Kb band. Transgenic mouse line ES5007 was derived from ES cell clone B12 (Table I). The remaining positive ES cell clones failed to establish germline transmission of the FAD mutation.

Initial miniSouthern-blot analyses identified five ES cell clones which appeared to contain Swedish/London FAD APP double mutation (data not shown). DNA extracted from pellets of expanded ES cell clones was examined by Southern-blot analysis using the restriction enzyme Xba I as described in Figure 14. This analysis confirmed that ES cell clones C82, C87, D25 and D92 contained the Swedish/London FAD m/hAPP double mutation while clones C52 and D49 did not. Transgenic mouse line ES5103 was derived from ES cell clone C87 (see Table I). The remaining positive ES cell clones failed to establish germline transmission of the FAD mutation. Confirmatory Southern-blot analyses identified multiple clones which carry the Swedish-FAD m/hAPP713 mutations (data not shown).

While identical in all other respects, the targeting vector encoding London-FAD m/hAPP does not carry the Xba I restriction site associated with the Swedish mutation. It was necessary, therefore, to identify restriction

enzymes which could distinguish between un-altered ES clones and those ES cell clones containing a targeted m/hAPP allele. The restriction enzymes Bcl I and BpM I were found to distinguish between DNA from a non-targeted ES cell clone (clone A1) and DNA from clone A21 which contains the Swedish-FAD m/hAPP gene locus (data not shown, see also figure 19). Bcl I and BpM I can be used to identify targeted clones derived from any of the aforementioned gene-targeting vectors.

Using restriction enzyme Bcl I, miniSouthern-blot analysis identified 6 ES cell clones which appeared to contain the London-FAD mutation (data not shown). Confirmatory Southern-blot analysis , using restriction enzyme BpM I, demonstrated that ES cell clones D12, D60, D70, D74, and D90 contained the London-FAD m/hAPP targeted locus while clone D45 did not (figure 18). After digestion with BpM I, three DNA fragments (~6 Kb, ~3.8 Kb, and ~2.2 Kb) are observed in non-targeted ES cells whereas targeted ES cell clones exhibited an additional ~4.8 Kb DNA fragment from the targeted allele (see A21 targeted for Swedish mutation. Transgenic mouse lines ES5401 and ES5403 were derived from ES cell clones D12 and D60 respectively (Table I). The remaining positive ES cell clones failed to establish germline transmission of the FAD mutation.

Germline-transmission of targeted m/hAPP genes

ES cells, confirmed to contain a targeted m/hAPP allele, were injected into the blastocoel cavity of a 3.5 day pre-implantation embryo (blastocyst). The injected blastocysts were then surgically reimplanted into pseudopregnant fosters and chimeras were born after approximately 17 days. The ES cells were derived from the 129SVEV inbred mouse strain which has a dominant agouti coat color gene. The blastocysts were derived from the C57BL/6 inbred mouse strain which carries a recessive black coat color gene. The coat color of chimeric mice whose cells are predominately derived from the ES

cells (designated as "high percentage chimeras") is mostly agouti with small patches of black. To establish germline transmission of the targeted APP gene, high percentage chimeric male mice were mated with either 129/SVEV inbreed or black Swiss outbreed females. The genotype of offspring from ES5007, ES5103, ES5401 and ES5403 breeding pairs was determined by either Southern-blot or PCR analyses.

The Southtern-blot analyses could distinguish between non-targeted, heterozygous, and homozygous progeny mice. The analyses utilized either Bcl I or BpM I restriction enzyme, and the ~3.0Kb intron 15 DNA fragment as probe (see Figure 14 for description of probe). A Southern-blot characterizing DNA from progeny of heterozygous ES5007 breeding pairs was performed. The technique can be applied to all the aforementioned transgenic lines. Bcl I digestion of non-transgenic (wt) mouse DNA and non-targeted ES cell DNA generated ~16 and ~8.5 Kb DNA fragments. However, Bcl I digestion of heterozygous transgenic mouse DNA and targeted ES cell DNA generated ~16 , ~8.5, and ~8.0 Kb DNA fragments. Digestion of homozygous mouse DNA with Bcl I liberated ~8.5 and ~8.0 Kb DNA fragments. BpM I digestion of non-transgenic (wt) mouse DNA and non-targeted ES cell DNA generated ~6.0, ~3.8, and ~2.2 Kb DNA fragments. Bpm I digestion of heterozygous transgenic mouse DNA and targeted ES cell DNA generated ~6.0 , ~4.8, ~3.8, and ~2.2 Kb DNA fragments Digestion of homozygous mouse DNA with BpM I liberated ~6.0 , ~4.8, and ~2.2Kb DNA fragments.

The genotype of offspring from ES5007, ES5103, ES5401 and ES5403 breeding pairs was also determined by PCR analyses using a combination of oligo pairs specific to human APP (H) and mouse APP (M) sequences. Like the Southern-blot technique, PCR analysis can distinguish between non-targeted, heterozygous, and homozygous progeny mice. As an internal standard, a 154 bp region of the mouse ribosomal subunit L32 gene is

amplified using the PCR oligo pair 6 and 7 (Figure 9). This control reaction was performed in each reaction. A 118 bp region specific to the mouse APP gene is amplified using the "M" oligo pair (oligos KC125 and KC132; Figure 9) and a 109 bp region specific to targeted m/hAPP gene in amplified using the "H" oligo pair (oligos KC125 and KC131; Figure 9). A PCR reaction using non-transgenic mouse DNA (wt) gives rise to a 118 bp fragment using the "M" oligo pair but no reaction product using the "H" oligo pair. Conversely, a PCR reaction using DNA from transgenic mice homozygous (homoz.) for the targeted m/hAPP gene gives rise to a 109 bp fragment using the "H" oligo pair but no reaction product is observed using the "M" oligo pair. A PCR reaction using DNA from heterozygous transgenic mice (heter.) gives rise to both mouse and human PCR reaction products.

Messenger RNA (mRNA) expression in transgenic mouse brain

Analysis of APP mRNA composition in control mouse and ES5007 mouse brain has been determined using both Northern-blot and rtPCR analyses. RNA analyses have yet to be performed on the ES5103, ES5401, and ES5403 lines (Table II).

APP mRNA transcripts from control and ES5007 mouse brain were detected by Northern-blot analysis using an approximately 900 bp Nru I- Xho I fragment from pMTI-2385B (human APP cDNA) as probe. Mouse beta-actin mRNA was detected using a 430 bp mouse beta-actin cDNA probe (430 bp PCR product generated using oligos KC137 and KC138; see Figure 9) and served as an internal standard. mRNA from human brain (Clonetech) served as a positive control.

mRNA from the Swedish-FAD m/hAPP gene was abundantly expressed in the brain from homozygous ES5007 mice. The amount of Swedish-FAD m/hAPP mRNA in ES5007 brain was determined by phosphoimage analysis and shown to be approximately 55% of the mAPP

mRNA levels observed in control mouse brain. In concordance, the APP mRNA levels in heterozygous ES5007 mouse brain were found be approximately 75% of the level observed in control mouse brain.

The reverse transcriptase-PCR (rtPCR) technique was used to identify mouse APP and m/hAPP transcripts in mouse brain. Homozygous ES5007 mice were found to express mRNA exclusively from the targeted Swedish-FAD m/hAPP gene. No mRNA species containing sequences from mouse APP exons 16, 17, or 18 was detected in homozygotes. Heterozygous ES5007 mice were found to express mRNA transcripts from both mouse APP and Swedish-FAD m/hAPP alleles.

mRNA was purified from control and transgenic mouse brain and cDNAs were prepared using reverse transcriptase and oligonucleotide RA49 as primer. A 367 bp DNA fragment was amplified from mouse APP and m/hAPP cDNA by PCR using oligonucleotides KC56 and RA49 (Figure 9). Oligonucleotides KC56 and RA49 exhibit sequence identity with both mouse and human sequences. The mouse and human sequences were distinguished from each other by the presence of a Sty I restriction site in the human cDNA and the absence of the Sty I site in the mouse cDNA. Digestion of the 367 bp PCR product from m/hAPP cDNA generates two fragments (151 bp and 216 bp) while the PCR product from the mouse APP cDNA is not digested and remains unchanged at 367 bp.

APP mRNA from control mouse brain was amplified by rtPCR to generate a 367 bp DNA fragment that was resistant to Sty I digestion. rtPCR amplification of m/hAPP mRNA from the brain of homozygous ES5007 mice gene generated two fragments (151 bp and 216 bp) upon digestion by Sty I. No 367 bp DNA fragment remained, demonstrating that mouse APP cDNA was not present. All three DNA fragments (151 bp, 216 bp, and 367 bp) were observed after Sty I digestion of rtPCR product derived from heterozygous ES5007 brain transcripts. As a control, Sty I digestion of PCR

products from human APP cDNAs derived from human mRNA and mRNA from a HEK293 cell line expressing human APP generated the 151 and 216 bp DNA fragments. As expected, Sty I failed to digest the 367 bp PCR product derived from mouse brain mRNA.

m/hAPP protein expression in transgenic mouse brain

Swedish-FAD m/hAPP protein is expressed in the brain of ES5007 mice. MAb 286.8 specifically immunoprecipitates human APP and but not mouse APP. The epitope for MAb 286.8 has been determined to lie within the N-terminus of the human beta-amyloid domain (P. Graham et al. 1994, Pharma Report MRC 00116). The m/hAPP gene product could be specifically immunoprecipitated from a ES5007 brain homogenate using the monoclonal antibody (MAb) 286.8. APP moieties were then visualized by Western-blot analysis using MAb 22C11 as the detection antibody. MAb 22C11can detect both mouse APP and m/hAPP. Therefore, if mouse APP was present after the immunoprecipitation it would have been detected by MAb 22C11. MAb 286.8 immunoprecipitated baculovirus derived human APP but did not recognize mouse APP in mouse brain homogenates.

The immunoprecipitations were performed using equal amounts of control mouse and ES5007 brain homogenates directly applied to the Western-blot. The relative intensities of the mouse APP and m/hAPP bands were equivalent.

Baculovirus derived human APP was directly applied to the Western-blot. An equal amount of human APP was detected after immunoprecipitation by MAb 286.8. It can be concluded, therefore, that MAb286.8 efficiently immunoprecipitated human APP.

The expression of Swedish-FAD m/hAPP protein was further demonstrated by Western-blot analyses using additional detection antibodies. m/hAPP was immunoprecipitated from a homogenate of ES5007 brain using

human-specific MAb286.8. APP was then detected by Western-blot analysis using either the polyclonal antibody (PAb) 369 or MAb 6E10 for detection. MAb 6E10 is human APP specific and recognizes the human beta-amyloid domain. Again MAb 286.8 immunoprecipitates human APP, Swedish-FAD m/hAPP but does not immunoprecipitate mouse APP.

Swedish-FAD m/hAPP protein is expressed in the brain of homozygous ES5007 mice at approximately 87% of the level observed for mouse APP in non-transgenic mice. The relative expression values were determined in 3 independent Western-blot analyses using homogenates of brains from 4 homozygous ES5007 and 4 non-transgenic mice. The levels of m/hAPP protein expression ranged from 62% to 130% of control mouse APP depending on the protocol. In one experiment, APP was immunoprecipitated from equal amounts of brain homogenates from non-transgenic and homozygous ES5007 mice using PAb 369. For the other two Western-blot analyses, MAb 4G8 was used to immunoprecipitate APP. In all Western-blots, mouse APP and Swedish-FAD m/hAPP were visualized using MAb 22C11 as the detection antibody.

<u>Processing of C-terminal domain of APP</u>

The Swedish FAD mutation significantly altered the proteolytic processing of the of APP resulting in a change in theC-terminal fragments of APP. The observed changes in processing was consistent with a predominat usage of the beta-secretase site over the alpha-secretase site.

Membrane preparations from brain homogenates were solubilized by detergents and APP holoprotein and C-terminal fragments were immunoprecipitated using MAb 4G8. Mouse APP and m/hAPP holoproteins were detected by Western-blot analysis using MAb 22C11. The C-terminal fragments from both mouse APP and m/hAPP were detected using PAb369 while C-terminal fragments derived exclusively from m/hAPP were detected

using human specific MAb 6E10. The expression level of m/hAPP in homozygous ES5007 (Swed-homoz) brain was found to be approximately 62% of the level observed for mouse APP.

Under normal conditions, the proteolytic processing of mouse APP resulted in the generation of 5 C-terminal fragments. This contrasts with the pattern observed with Swedish-FAD m/hAPP where only the two largest C-terminal fragments were observed. The second largest C-terminal fragment (fragment 2) co-migrated with the LEC100 standard. The electrophoretic mobility of LEC100 was expected to closely resemble that of the C-terminal fragment released after the cleavage by beta-secretase. LEC100 consists of amino acid residues Leu, Gly, and Met juxtaposed with the beta-amyloid, transmembrane, and cytoplasmic domains of APP. spLEC100 (sp designates APP signal peptide, see below) was stably expressed in HEK293 cells (obtained from Sandra Reuter), a membrane homogenate prepared, and an aliquot was applied to the gel. In cells, LEC100 is generated after the signal peptide (sp) is proteolytically removed from spLEC100 during protein translation.

For other aspects of the nucleic acids, polypeptides, antibodies, etc., reference is made to standard textbooks of molecular biology, protein science, and immunology. See, e.g., Davis et al. (1986), *Basic Methods in Molecular Biology*, Elsevir Sciences Publishing, Inc., New York; Hames et al. (1985), *Nucleic Acid Hybridization*, IL Press, *Molecular Cloning*, Sambrook et al.; *Current Protocols in Molecular Biology*, Edited by F.M. Ausubel et al., John Wiley & Sons, Inc; *Current Protocols in Human Genetics*, Edited by Nicholas C. Dracopoli et al., John Wiley & Sons, Inc.; *Current Protocols in Protein Science*; Edited by John E. Coligan et al., John Wiley & Sons, Inc.; *Current Protocols in Immunology*; Edited by John E. Coligan et al., John Wiley & Sons, Inc.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The preceding preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The entire disclosure of all applications, patents and publications, cited below are hereby incorporated by reference.

From the foregoing description, one skilled in the art can easily ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

Tables

Table I

Transgenic line	ES cell clone	Targeting vector	Gene Product	mAPP mutations
ES5007	B12	pMTI-2398	Swedish-FAD m/h APP	KM(670,671)NL; G(676)R; F(681)T, R(684)H
ES5103	C87	pMTI-2454	Swedish/London-FAD m/h APP	KM(670,671)NL; V(717)I; G(676)R, F(681)T, R(684)H
ES5215	A54	pMTI-2455	Swedish-FAD m/hAPP713	KM(670,671)NL; T(714)stop; G(676)R, F(681)T, R(684)H
ES5401	D12	pMTI-2453	London-FAD m/h APP	V(717)I; G(676)R, F(681)T, R(684)H
ES5403	D60	pMTI-2453	London-FAD m/h APP	V(717)I; G(676)R, F(681)T, R(684)H

Table II

Transgenic Line	Germline-Transmission	m/h APP mRNA	m/h APP Protein	Altered C-Terminal Processing
ES5007	yes	yes	yes	yes
ES5103	yes	n.d.	yes	yes
ES5401	yes	n.d.	yes	n.d.
ES5403	yes	n.d.	yes	n.d.

n.d.: not determined

G.R. Askew, T. Doetschman and J.B. Lingrel, *Mol. Cell Biol.*, 13:4115-24 (1993)

C. Bonnerot, G. Grimber, P. Briand and J.F. Nicolas, *Proc. Natl. Acad. Sci. USA*, 87:6331-6335 (1990)

A. Bradley, in E.J. Robertson (ed.), *Production and analysis of chimeric mice*, IRL Press, Oxford and Washington, DC, pp. 113-151 (1987)

R.L. Brinster, J.M. Allen, R.R. Behringer, R.E. Gelinas and R.D. Palmiter, *Proc. Natl. Acad. Sci. USA*, 85:836-840 (1988)

X.D. Cai, T.E. Golde and S.G. Younkin, *Science*, 259:514-516 (1993)

M. Citron, C. Vigo-Pelfrey, D.B. Teplow, C. Miller, D. Schenk, J. Johnston, B. Wilblad, N. Venizelos, L. Lannfelt and D.J. Selkoe, *Proc. Natl. Acad. Sci. USA*, 91:11993-11997 (1994)

T. Dyrks, E. Dyrks, C.L. Masters and K. Beyreuther, *FEBS Lett.*, 324:231-236 (1993)

D. Games, D. Adams, R. Alessandrini, R. Barbour, P. Berthelette, C. Blackwell, T. Carr, J. Clemens, T. Donaldson, F. Gillespie, et al., *Nature*, 373:523-527 (1995)

S.A.e.a. Gravina (1995)

F. Grosveld, G.B. van Assendelft, D.R. Greaves and G. Kollias, *Cell*, 51:975-985 (1987)

P. Hasty, R. Ramirez-Solis, R. Krumlauf and A. Bradley, *Nature*, 350:243-246 (1991)

L.S. Higgins et al. (1995)

L.S. Higgins, D.M. Holtzman, J. Rabni, W.C. Mobley and B. Cordell, *Ann. Neurol.*, 35:598-607 (1994)

J.Kang and B. Muller-Hill, *Biochem. Biophys. Res. Commun.*, 166:1192-1200 (1990)

F.M. La Ferla, B.T. Tinkle, C.J. Bieberich, C.C. Haudenschild and G. Jay, *Nat. Genet.*, 9:21-30 (1995)

F.M. La Ferla, D.A. Kappel Hall, L. Ngo and G. Jay, In manuscript (1995)

B.T. Lamb, *Nat. Genet.*, 9:4-6 (1995)

B.T. Lamb, S.S. Sisodia, A.M. Lawler, H.H. Slunt, C.A. Kitt, W.G. Kearns, P.L. Pearson, D.L. Price and J.D. Gearhart, *Nat. Genet.*, 5:22-30 (1993)

H.G. Lemaire, J.M. Salbaum, G. Multhaup, J. Kang, R.M. Bayney, A. Unterbeck, K. Beyreuther and B. Muller-Hill, *Nucleic Acids Res.*, 17:517-522 (1989)

B.E. Pearson and T.K. Choi, *Proc. Natl. Acad. Sci. USA*, 90:10578-10582 (1993)

E.J. Robertson, in E.J. Robertson (ed.), *Embryo-derived stem cell lines*, IRL Press, Oxford and Washinton, DC, pp. 71-112 (1987)

M. Rubinstein, M.A. Japon and M.J. Low, *Nucleic Acids Res.*, 21:2613-2617 (1993)

W. S. Simonet, N. Bucay, S.J. Lauer and J.M. Taylor, *J. Biol. Chem.*, 268:8221-8229 (1993)

A. Stacey, A. Schnieke, J. McWhir, J. Cooper, A. Colman and D.W. Melton, *Mol. Cell Biol.*, 14:1009-1016 (1994)

N. Suzuki, T.T. Cheung, X.D. Cai, A. Odaka, L. Otvos, Jr., C. Eckman, T.E. Golde and S.G. Younkin, *Science*, 264:1336-1340 (1994)